
BIOINFORMATICS ANALYSIS OF EPIGENETIC VARIANTS ASSOCIATED WITH MELANOMA

Katarzyna MURAT

*submitted for the degree
of Master of Philosophy*

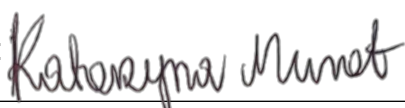
Department of Chemistry and Biosciences
University of Bradford

2018

Declaration of Authorship

I, Katarzyna MURAT, declare that this thesis titled, "BIOINFORMATICS ANALYSIS OF EPIGENETIC VARIANTS ASSOCIATED WITH MELANOMA" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: 

Date: 30.07.2018

Abstract

BIOINFORMATICS ANALYSIS OF EPIGENETIC VARIANTS ASSOCIATED WITH MELANOMA

by Katarzyna MURAT

The field of cancer genomics is currently being enhanced by the power of Epigenome-wide association studies (EWAS). Over the last couple of years comprehensive sequence data sets have been generated, allowing analysis of genome-wide activity in cohorts of different individuals to be increasingly available. Finding associations between epigenetic variation and phenotype is one of the biggest challenges in biomedical research. Laboratories lacking dedicated resources and programming experience require bioinformatics expertise which can be prohibitively costly and time-consuming. To address this, we have developed a collection of freely available Galaxy tools (Poterlowicz, 2018a), combining analytical methods into a range of convenient analysis pipelines with graphical user-friendly interface. The tool suite includes methods for data preprocessing, quality assessment and differentially methylated region and position discovery. The aim of this project was to make EWAS analysis flexible and accessible to everyone and compatible with routine clinical and biological use. This is exemplified by my work undertaken by integrating DNA methylation profiles of melanoma patients (at baseline and mitogen-activated protein kinase inhibitor MAPKi treatment) to identify novel epigenetic switches responsible for tumour resistance to therapy (Hugo et al., 2015). Configuration files are publicly published on our GitHub repository (Poterlowicz, 2018b) with scripts and dependency settings also available to download and install via Galaxy test toolshed (Poterlowicz, 2018a). Results and experiences using this framework demonstrate the potential for Galaxy to be a bioinformatics solution for multi-omics cancer biomarker discovery tool.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor Dr Krzysztof Poterlowicz for the continuous support of my MPhil study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my study. Besides my advisor, I would like to thank colleagues, for their insightful comments and encouragement. Last but not the least, I would like to thank my family for supporting me spiritually throughout writing this thesis and my life in general.

Contents

Declaration of Authorship	i
Abstract	ii
Acknowledgements	iii
1 Introduction	1
1.1 Skin Cutaneous Melanoma	1
1.2 Epigenetics Regulation of Disorders	7
1.3 Computational Epigenetics	11
1.4 Public Databases for Epigenetic Information	15
2 Computational Methods for Epigenetics	18
2.1 Epigenome-Wide DNA Methylation Profiling	18
2.1.1 Pre-processing	20
2.1.2 Mapping and Graphical Visualization	21
2.1.3 DMR and DMP Identification	22
2.1.4 Peak Detection	23
2.2 Galaxy Tools Development	24
3 Development of Galaxy Tools	28
3.1 EWAS - Galaxy Tool Suite	28
4 Application of the Tools	34
4.1 Enhancement for the Computational Methods	34
4.2 Clinical Relevance Validation	35
4.3 Discovery and Validation of a Predicted SKCM Epigenetics Variants	37

A Galaxy EWAS Training	40
B Table of Differentially Methylated Regions	45
C Table of DMR's Annotation	58
D Table of DMR's Functional Annotation (GO)	79
Bibliography	84

List of Figures

1.1	The layers of human skin (epidermis, dermis and hypodermis), as well as an inset with a close-up view of the types of cells in the skin (keratinocyte both squamous cells and basal cells, and melanocytes) (Gordon, 2013).	2
1.2	Galaxy Project (public) interface consisting of tool search panel, working space area and history.	15
2.1	The Infinium Methylation Assay Technology (Illumina, 2018) . .	20
2.2	Galaxy's graphical workflow editor, show part of a sample workflow.	26
2.3	Local Galaxy Instance running via command line with custom interface consisting predefined tools sections, tailored working space area and test history.	27
3.1	Simplified workflow for analysing epigenetics data	29
3.2	Screenshot from the Galaxy Workflow Editor, showing EWAS example workflow discussed in the Analyses section.	33
4.1	Quality Control Plot representation of melanoma pre and post MAPKi treatment samples.	36
4.2	Functional Annotation of DMR's found in melanoma biopsies pre and post MAPKi treatment.	38
A.1	Data Upload minfi_read450k Tool Interface	41
A.2	Quality Assessment (minfi_qc) Tool Interface	42
A.3	minfi_dmr Tool Interface	43
A.4	DMR Training Track launch UCSC Genome Browser	44

List of Tables

3.1	Summary of the EWAS suite tools inputs and outputs	30
-----	--	----

List of Abbreviations

EWAS	Epigenome-wide association study
TSG	Tumor Suppressor Genes
NGS	Next-generation sequencing
GEO	Gene Expression Omnibus
SNP	Single-nucleotide polymorphism
DMR	Differentially Methylated Regions
DMP	Differentially Methylated Positions
MAPK	Mitogen-activated protein kinase
MITF	Microphthalmia Associated Transcription Factor
TCGA	The Cancer Genome Atlas
ENCODE	Encyclopedia of DNA Elements
SDK	Software Development Kit
API	Application Programming Interface
FDR	False Discovery Rate

*Dedicated to my beloved Parents Monika and
Kazimierz Murat*

Chapter 1

Introduction

1.1 Skin Cutaneous Melanoma

To better understand skin cancer, we need a basic knowledge of the skin itself. Human skin is considered to be an extraordinary organ providing the main barrier between the internal and external environment. This barrier is designed to protect the human body against a large number of environmental stressors such as temperature, microbial pathogens, chemical agents and UV (Tobin, 2006). Normal skin consists of layers of the epidermis, papillary and reticular dermis, and hypodermis (subcutaneous fat layer) shown in Figure 1.1 below. The epidermis is composed of three living cell layers which from the epidermis – dermis junction include *stratum basale*, *s. spinosum*, and *s. granulosum* followed distally by the non-viable but biochemically active *stratum corneum*. The latter is composed of dead keratinizing or cornifying stratified epithelium cells that have migrated outward from the basal layer (Tobin, 2006). The *stratum basale* (basal layer) is the deepest sublayer, containing three major type of cells such as predominant keratinocytes, and low numbers of melanocytes, and Merkel cells (Tobin, 2006). Underlying the epidermis is the dermis, which provides support and nutrients for the epidermis. The dermis is composed of collagen, reticulin and elastic fibers, and houses the appendages including hair follicles and sweat glands (Tobin, 2006). It also contains vascular, neural and lymphatic systems with multiple receptors for touch, temperature and pain. In the human skin, melanocytes reside in the *stratum*

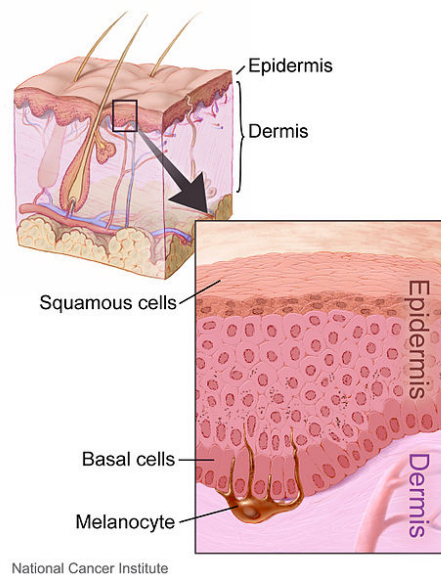


Figure 1.1: The layers of human skin (epidermis, dermis and hypodermis), as well as an inset with a close-up view of the types of cells in the skin (keratinocyte both squamous cells and basal cells, and melanocytes) (Gordon, 2013).

basale as shown on Figure 1.1 and produce the pigment melanin, which protects the skin from ultraviolet radiation (UVR) (Tobin, 2006). Early stages of melanocyte development initiates with the fate commitment of precursor cell lineages from the neural crest, which is then followed by cell migration and niche localization. Many genes involved in melanocyte development and pigmentation have been characterized as key to identifying the genes and proteins involved in melanoma skin cancer. Transformed melanocytes that develop into melanomas increase quickly and spread aggressively to other parts of the body making it one of the most deadly cancers (Uong and Zon, 2010). Melanocytes arise from, a certain class of stem cells called neural crest cells (NCCs) (Erickson and Reedy, 1998). The neural crest is induced at the time of gastrulation, in the zone between the neural and non-neural ectoderm (Erickson and Reedy, 1998) and gives rise to a number of cells populations. Pigment cells are generated from multipotent neural-melanocytic progenitors or bipotent glial-melanocytic precursors melanoblast (Dupin and Le Douarin, 2003). Melanoblast cells migrate from the trunk neural crest cells dorsolaterally between the ectoderm and dorsal surface compartments (Dupin and Le

Douarin, 2003). Melanoblasts are exposed to a range of developmental cues that regulate their specification process (Dorsky, Moon, and Raible, 1998). This developmental process is governed primarily by key signaling molecules including Wnt and BMP (Kléber et al., 2005), and other complex pathways. Wnt protein mediates a fate switch between melanogenesis and gliogenesis through activated beta-catenin, as overexpression of beta-catenin leads to an increase of melanocytes and a loss of glial derivatives (Dorsky, Moon, and Raible, 1998). On the other hand BMP signaling suppresses neural crest cell differentiation into melanocytes and appear to have opposing effects to Wnt (Dhara and Stice, 2008). Their coordinated expression is required for the normal melanocyte growth and development (Jin et al., 2001). A range of experiments identified a series of interactions among molecules relevant for melanocyte development (Giebel and Spritz, 1991). These studies have elucidated genetic pathways such as transcription factor (*MITF*) and tyrosine kinase (*KIT*) genes with essential functions for cell homeostasis including survival, cell cycle and metabolism. *KIT* and *MITF* show complex interactions, in that *MITF* is needed for the maintenance of *KIT* expression in melanoblasts and *KIT* signaling modulates *MITF* activity and stability in melanocyte cell lines (Kawakami and Fisher, 2017). In melanocyte development *MITF* stimulates melanogenesis by activating some key pigmentation-related genes transcription activation (e.g. TYR, TYRP1, MLANA) and anti-apoptotic genes up-regulation (e.g. BCL2, BCL2A1) (Garraway et al., 2005). On the other hand *KIT* plays a crucial role in the survival, migration, and development of the melanocyte lineage (Giebel and Spritz, 1991). This effect on melanin production may explain why disruption of those genes lead to defects in melanocyte numbers causing variety of skin disorders including cancer. Many genes involved in melanocyte development have also been implicated in melanoma progression. Several genetic pathways regulate cell growth and survival and play important roles in normal embryonic development as well as melanoma progression (Uong and Zon, 2010).

The mitogen activated protein kinase (MAPK) is a type of protein kinase

that is specific to the amino acids serine and threonine. MAPK signaling pathway also known as the Ras-Raf-MEK-ERK pathway is one of several potentially target-able pathways in melanoma. Mutations that over-activate BRAF and NRAS are found in nearly 65% and 20% of melanomas respectively (Gilbertson et al., 2006). Understanding how melanomas acquire resistance to BRAF and NRAS inhibitors via genetic alterations is important, as leads to the reactivation of the MAPK pathway, and so is a focus of drug discovery potential (Hugo et al., 2015). Recent studies have provided further insights of many genes that are up- and down-regulated in melanoma development. For example, *MITF* and *KIT* have a crucial role in the melanocyte life cycle (Garraway et al., 2005). Over-expression of these genes can caused inappropriate cell cycle progression and as a result to this changes development of the tumour (Garraway et al., 2005). Cutaneous melanomas incidence is rapidly increasing and is projected to rise by 7% in the UK between 2014 and 2035, reaching 32 cases per 100,000 people by 2035 (Smittenaar et al., 2016). The most important factors suggested for this increased risk include unprotected exposure to intense natural ultraviolet light (sun rays) and artificial ultraviolet radiation (e.g. tanning beds, sunbeds), permanent mechanical or chemical irritation and genetic predisposition (e.g. red-hair phenotype (MC1R variants) and familial atypical mole syndrome or family of atypical features)(Ghiorzo et al., 1999). Suspected skin melanoma may induce changes that have developed *de novo* or on the basis of a pigmented lesion (thickening, surface change, discolouration and margins or occurrence of itching and / or bleeding)(Ghiorzo et al., 1999). The early identification and location of the primary tumour creates a unique chance of curing skin melanoma. Prognosis appears in part to be a function of tumour size and stage of invasion with tumour thickness (Breslow, 1970). Physical examination should include questions about the condition of the skin (i.e. information about changes in existing skin lesions or the occurrence of new lesions) and factors that increase the risk of skin melanomas (e.g. sunburn, use of solarium, occurrence of melanomas in

the family). The most important element allowing early diagnosis is the examination of the skin, which should be performed by the doctor during the patient's visit to the outpatient clinic or during hospitalization. The principle is to assess the skin of the whole body in good lighting, checking all the skin areas especially ones that are difficult to access (head, feet, inter-digital spaces, genital and anal areas). In low-risk melanomas (pT1a) the melanoma is less than 1mm thick, not ulcerated and no other investigations are necessary (Dummer et al., 2012). Computed tomography (CT) and positron emission tomography-computed tomography (PET-TK) should be considered in patients with higher tumour stages (pT1b–pT3a) especially in the presence of clinical metastases to lymph nodes or isolated metastases to distant organs (Dummer et al., 2012). In the case of clinical metastases to the inguinal lymph nodes, a CT scan or magnetic resonance imaging (MRI) of the pelvis are recommended before surgical treatment and sentinel node biopsy (Dummer et al., 2012). The identification of clinical and pathomorphological prognostic features is aimed at understanding the biology of cancer and facilitating the planning of appropriate treatment for an individual patient, including the risk of relapse and the likelihood of survival after treatment (Dummer et al., 2012). The risk factors that achieved convincing or highly suggestive evidence regarding contribution to the development of skin cancer are genetic predisposition, sun exposure, skin colour, photo-sensitivity and age. Cases of cancer in family members increase the risk of getting cancer caused by the transmission of mutations associated with pigment cells development (Uong and Zon, 2010). Nevertheless, knowing genes responsible for melanoma occurring in families contributed significantly to understanding the molecular mechanisms of this disease (Gordon, 2013). Observation of the patients whose relatives have suffered from the disease, gave the first insight into the model of melanoma progression, which can be accelerated on the basis of congenital disorder (Gordon, 2013). In the case of about 50% of families tested mutations were found in two closely related genes CDKN2A and CDK4. CDKN2A encodes two suppressor genes p16INK4a and p14ARF inhibiting progress

of the cell cycle (Ghiorzo et al., 1999). A close relationship was established between the genetic diversity at the melanocortin receptor 1 (MC1R) and pigmentation of the skin and hair, and recently also a predisposition to developing melanoma (Ghiorzo et al., 1999). Inheriting CDKN2A and CDK4 gene mutations does not cause cancer by themselves. Skin DNA damage caused by UV rays leads to numerous acquired or somatic genetic changes. Multiple research studies show a correlation between increased risk of skin cutaneous melanoma (SKCM) and sun exposure (Ghiorzo et al., 1999). Moreover, It was explained that the majority of melanoma occur on skin occasionally exposed (e.g. abdomen skin) to sunlight when patients are young age and this tendency decreases with age of exposure. On the other hand in elderly patients melanoma tend to arise in skin constantly exposed to sun (e.g. face, neck, arms) (Ghiorzo et al., 1999). There is no sure way to prevent melanoma, as age, gender, race, and family history can't be controlled. However, regular skin checks and limiting unprotected UV exposure can help (Ghiorzo et al., 1999). The primary prevention of melanoma is concerned with a reduction in the risk factors for skin cancer, most notably sun exposure and sunburn. The well-proven methods are off covering the skin, wear a hat and, applying significant amounts of sunscreen are still are not generally well adhered to, and so rates remain very high and increasing (Gordon, 2013). Identification of clinical and pathomorphological features of prognosis aims to understand the biology of cancer and facilitating the planning of proper treatment for individual patients, including the risk of disease recurrence and probability of survival after treatment. The most important prognostic factors in melanoma patients without the presence of metastases is vertical tumour thickness (Breslow's depth), presence of histologically recognised ulceration, mitotic rate and level of invasion (Clark's level) (Gospodarowicz and Wittekind, 2017). The staging system most often used for melanoma is the American Joint Committee on Cancer (AJCC) TNM system, which is based on 3 key pieces of information: tumor thickness (T), spread to nearby lymph nodes (N) and metastasis

to distant sites (M). The earliest clinical stage of melanomas is stage 0 (carcinoma in situ), and then range from stages I through IV. As a rule, the lower the number, the less the cancer has spread. A higher number, such as stage IV, means cancer has metastasized to distant organs (Gospodarowicz and Wittekind, 2017). The majority of patients with stage 0 require surgical excision only but more advanced cancers often require other more complex treatments including immunotherapy. Immunotherapy is the use of medicines to stimulate immune system so that the latter can recognize and destroy cancer cells more effectively. Among stage III and IV tumours with a poor prognosis and high risk of recurrence it is recommended to perform an examination of their BRAF gene. The identification of the BRAFV600E mutation will drive the appropriate BRAF inhibitor treatment strategy (Carbognin et al., 2015). Independently of the status of the BRAF mutation, the immune system can be stimulated with antibodies anti-PD-1 antibodies (e.g. nivolumab or pembrolizumab) or by drugs like ipilimumab (anti-CTLA4 antibody). The sequence of treatment (especially in the presence of mutations BRAF) currently is still not specified. The combination of BRAF and MEK inhibitors is associated with a high response rate (approximately 70%) and a rapid improvement in the symptoms of the disease while treatment with anti-PD-1 antibodies brings less response rates, but they are more long-lasting (Carbognin et al., 2015). Detection of genomic alterations in the melanoma can help identify patients who may benefit from an experimental approach in clinical trials and future treatment design.

1.2 Epigenetics Regulation of Disorders

The first reports on epigenetics and the mechanisms of epigenetic modifications were initially published in the 1950s (Holliday, 2006). The number of epigenetics-related studies published since then is immense and rapidly growing (July 2018 over 63 thousand records founded in PubMed database of (Biotechnology Information, 2018), which indicates great interest in this

topic. New approaches, including DNA methylation and histone modifications are very important in various biological processes such as transcription (Busslinger, Hurst, and Flavell, 1983), genomic imprinting (Reik et al., 1987), developmental regulation (Antequera, Macleod, and Bird, 1989), mutagenesis (Cooper and Youssoufian, 1988), transposition (Banks and Fedoroff, 1989), DNA repair (Brown and Jiricny, 1988), X chromosome inactivation (Pfeifer et al., 1990), chromatin organization (Lewis and Bird, 1991). The renewed research in epigenetics has led to novel findings about aberrations in cytosine-methylation in the pathogenesis of neoplastic (Niller, Wolf, and Minarovits, 2009), neurodegenerative (Jakovcevski and Akbarian, 2012), psychiatric diseases (Abdolmaleky, Thiagalingam, and Wilcox, 2005), various cancers (Dawson and Kouzarides, 2012), immune disorders and pediatric disorders (Uddin et al., 2010). The term epigenome means "on top of" genome and refers to specific changes in genome regulatory activity occurring in response to environmental stimuli (Egger et al., 2004). Epigenetic modifications do not change the underlying DNA sequence, but can cause multiple changes in cellular function and gene expression (Egger et al., 2004). However, those changes still remain unknown in many progressive and incurable diseases such as cancers including one of the deadliest, melanoma (Esteller and Herman, 2002). A modification very widely studied currently is DNA methylation. The methylation of cytosine at the 5-position (m5C) in eukaryotic DNA is the only modification present in the genomes of all vertebrates and flowering plants (Bird, 1986). The role of this mark is so important that many researchers consider m5C to be the 5th base of DNA. The analysis of the human methylation profile showed that methylation typically occurs at cytosine in CpG dinucleotides (Bird, 1986). CpG's are not evenly distributed in the genome, with the majority of them grouped into non-methylated sequences islands close to promoter regions (Bird, 1987). Methylation of CpG islands associated with gene regulatory regions can cause genomic instability and lead to the development of many progressive and incurable diseases including cancer (Sandoval et al., 2011). The field of epigenetics is

quickly growing and with it the understanding that both the individual's environmental exposure and personal lifestyle can interact with the genome itself. Research indicates that most human diseases manifest from the interaction of genetic variants caused by the influence of epigenetic changes (Holliday, 2006). These changes may be reflected at various stages of development or even in later generations (Holliday, 2006). For example, a mother's exposure to pollution (before pregnancy) could impact her child's asthma or a father's diet could define predisposition to obesity. Genetic variants discovery can motivate society to provide healthy lifestyle and preventional protection against harmful environmental impact (Egger et al., 2004). Additionally, if the epigenetic signatures of chemical/pollution exposures can be identified than they can be limited to the minimum and the risk of epigenetic changes will decrease (Egger et al., 2004). Clinical applications of epigenetic changes, especially epigenetic functional regulation of gene expression has become one of the most important research topics in recent years, including in relation to tumour pathogenesis (Esteller and Herman, 2002). The main reason for this interest was the finding that methylation of chromatin and DNA methylation in the transcriptional silencing of genes related to oncogenesis (Esteller and Herman, 2002). In 1983 an experiment performed by Feinberg and Vogelstein found that genes of colorectal cancer cells obtained from primary human tumour tissues were substantially hypomethylated compared with normal tissues (Feinberg and Vogelstein, 1983). Nowadays, epigenetic changes discovery can be used as clinical biomarkers for early cancer molecular diagnosis, or as a treatment in gene targeted therapy in later stages (Esteller and Herman, 2002). There is also evidence showing that abnormal DNA methylation has been observed in immune disorders (Meda et al., 2011). Studies indicate that environmental agents, including stress and diet, combine to inhibit T-cell DNA methylation and epigenetic alterations, for example in lupus-like autoimmunity (Rakyan et al., 2011). However, epigenetics is such a new field of science that in most cases, its impact has not been fully demonstrated especially in terms of neurodegenerative disorders, as we still don't know much

about them (Jakovcevski and Akbarian, 2012). Several reports have associated Alzheimer's disease with difference in DNA methylation patterns related to brain function and ageing (Egger et al., 2004). Epigenetic errors appear one by one and findings from autopsies of brain tissue from patients with neurodegenerative disorders shed a new light on therapy in terms of incurable cases (Egger et al., 2004). Clinical applications of epigenetics, combined with the rise of technology, will lead to molecular diagnosis for which targeted treatments can be developed. (Bock and Lengauer, 2008).

DNA Methylation in Melanoma

There have been several investigations on the relationship between DNA methylation and melanoma. Genome-wide studies of DNA methylation have been performed and the changes in methylation level have been identified as potentially important in melanoma development, progression, and metastasis (Sager, 1989). DNA hyper-methylation of CpG islands at promoter sites has been widely described as a indicator of tumorigenesis by silencing tumor suppressor genes (TSG) (Sager, 1989). In 1995, Whelan et al. demonstrated that loss of function of the TSG gene cyclin-dependent-kinase inhibitor 2 (CDKN2) can lead to tumorigenesis (Whelan, 1995). To date, several studies have revealed association of TSG to melanoma. This is evident in the case of inactivation of RAS Association Domain Family Protein 1 (RASSF1A) which can be found in significant number of melanomas (and in around 90% of melanoma cell lines) (Maat et al., 2007) Moreover, BRCA1 associated protein1 (BAP1) silencing somatic mutations can be seen in more than 80% of metastasising tumours (Harbour, Onken, and Council, 2010). Apart from CDKN2, RASSF1A and BRCA1, which have been intensively discussed in multiple cancers, hypermethylation of other genes has also been associated with melanoma. The HOX gene family, a member of the homeodomain-containing transcription factors, has been used as a biomarker in various human cancers (Pramio et al., 2017) and are also regulated at the nuclear-cytoplasmic transport level in skin cancers (Pramio et al., 2017). Thus, the differential methylation of genes

in melanoma tumors provides an opportunity to more fully understand development of this cancer and to develop new therapeutic agents (Pramio et al., 2017).

MAPKi Targeted Therapy for Melanoma

Pathway targeted therapies (BRAF/MEK inhibitors) and immune checkpoint inhibitors have revolutionised melanoma treatment. Approximately 40% of melanomas express carcinogen BRAF mutations that is fundamental for the mitogen-activated protein kinase (MAPK) activation (Flaherty, 2012). Early studies showed that BRAF gene mutation require MAPK signalling for their survival (Flaherty, 2012). Over the past decade researchers have performed multiple clinical trials and experiments based on agents which target these pathways. In one of the largest trials on BRAF mutated metastatic melanoma researchers obtained over 422 samples from confirmed cutaneous melanoma stage III and IV BRAF mutation positive patients (Hauschild et al., 2012) subjected to MAPKi targeted therapy. Patients were treated with dabrafenib inhibitor of the associated enzyme B-Raf, which plays a master role in the regulation of cell development (Hauschild et al., 2012). While resistant tumours showed a reactivated MAPK pathway, there was a significant treatment effect (over 50%) compared with chemotherapy treatment (Hauschild et al., 2012). Results like these have helped to direct future strategies towards MAPKi therapy. However, melanoma resistant to MAPKi inhibitors still requires additional trials and tests to discover mechanisms responsible for intratumoral immunity (Hugo et al., 2015).

1.3 Computational Epigenetics

Recently, various experimental techniques have been developed for genome-wide mapping of epigenetic information. Next-generation sequencing (NGS) was shown to be useful for DNA methylation and histone modification profile detection (Harris et al., 2010). However, the latter study failed to consider

the differing categories of epigenetic data distribution, complexity level and data quality issue. Bioinformatics analysis and interpretation of epigenetic records have become one of the major challenges of recent years (Gautam et al., 2018). Working with genomic data requires some knowledge of computer languages like C++, S, R, Perl, Python, and many others were used to write genomic libraries and packages. While specific tools for data conversion and handling have been developed, unfortunately user friendly tools are still a rarity. The algorithms come first, with the interface following later. Powerful and flexible analysis systems do exist but they require computing and statistical knowledge. R is a widely used programming language and free software environment for statistical computing and graphics that is supported by the R Foundation for Statistical Computing (Matloff, 2011). It is a GNU operating system project that runs on different operation systems including Microsoft Windows, macOS, and UNIX, and is administered by CRAN project network (Matloff, 2011). R is an implementation of the S statistics language and quickly became more popular than S itself. Firstly due to the fact it is freely available and secondly, the involvement of more users means more developers and a wider contribution network. It also gives the possibility to use it in a more user-friendly way via graphical user interfaces (GUIs) such as RStudio (Racine, 2012), Deducer (Fellows, 2012), and Rattle (Williams, 2009). R is a very lean and functional language which allows the user to divide complex processes into modules (packages) (Paradis, Claude, and Strimmer, 2004) that can be customized depending on the user requirements and accessed through the Bioconductor website (www.bioconductor.org) (Gentleman et al., 2004). Bioconductor is a free, open source and open development R-based software project for the analysis and comprehension of high-throughput genomic data (Gentleman et al., 2004). To date, software development taken by the Bioconductor project provides a range of resources: a web based repository, nearly 1560 software packages, hundreds of metadata packages and a number of experimental data packages, publications, slides and training materials are constantly improved by users and developers (Gentleman et

al., 2004). Although R and Bioconductor present many advantages as they are a flexible and reproducible solution, they still require users to know the statistical and programming concepts according to the analysis to avoid erroneous results (Matloff, 2011). The major advance offered by next-generation sequencing (NGS) technologies is the ability to produce, in some cases, in excess of one billion short reads per instrument run, which makes them useful for many biological applications. A Bioconductor software (Robinson, McCarthy, and Smyth, 2010) offers a suite of tools for analyzing and visualizing a variety of sequencing approaches based on peak-finding algorithms, data summarisation and visualisation (Bock and Lengauer, 2008). Combining chromatin immunoprecipitation assays (ChIP) can be analysed with multiple peak callers e.g. MOSAiCS (Sun et al., 2013), ChIPseqR (Humburg et al., 2011), BayesPeak (Spyrou et al., 2009). Differential expression analysis of RNA-seq data can be run with EdgeR (Robinson, McCarthy, and Smyth, 2010) or DESeq2 (Love, Huber, and Anders, 2014) solution. The unique data distribution characteristics of DNA methylation require the development of dedicated bioinformatics and computational tools. Bioconductor packages provide analysis of multiple DNA methylation methods such as bisulfite sequencing (methylPipe (Kishore et al., 2015), DMRcate (Peters et al., 2014)) and one of the most common techniques Infinium Methylation Assay 450k (Dedeurwaerder et al., 2011) (DMRcate (Peters et al., 2014), Minfi (Hansen and Aryee, 2012), ChAMP (Morris et al., 2013), methylumi (Davis et al., 2012), RnBeads (Assenov et al., 2014)). The goal of the analysis is mainly to identify differentially methylated regions (DMRs) or differentially methylated positions (DMPs) integrate, and then visualize epigenomic data sets (Bock and Lengauer, 2008). The last but not least step of each of the above is enrichment analysis to discover functions (Gene Ontology (Ashburner et al., 2000)) and pathways (KEGG (Kanehisa and Goto, 2000)). It is still a case of manually inputting data as gene list into the DAVID (Dennis et al., 2003) functional annotation tool or running the tool via command line.

Application of Galaxy for Epigenetic Research

Developers aims to make computational biology accessible to everyone. Thus, they have devised Galaxy an open, web-based platform for accessible, reproducible, and transparent data-intensive research (<https://usegalaxy.org>) see Figure 1.2 (Goecks, Nekrutenko, and Taylor, 2010). This platform is accessible for users of every level of interest and knowledge. It features hundreds of tools and workflows, which are easily to run via user friendly interface (Goecks, Nekrutenko, and Taylor, 2010). Reproducible workflows and interactive histories allow users to repeat tasks and fully understand the analysis. This followed by transparent publication of data processing and sharing these on-line creates an opportunity to verify results and methods by the community. The main Galaxy server has more than 124,000 registered users worldwide, who run approximately 245,000 analysis jobs each month (Goecks, Nekrutenko, and Taylor, 2010). Newly registered users can address their needs and get support from more advanced developers, administrators or educators involved in the project. I outline in Figure 1.2 a scheme that in three panels illustrate the available tools section, working space and history on Galaxy, containing inputs and outputs of the analysis. Analysis tools of interest can be searched by name or by the topic/category they belong. When a tool is selected, it is shown in the working space ready to use. When a user apply inputs and custom settings then only one click separates him from getting results. Output as an inputs datasets are added to the history panel and they can be modified, renamed, shared or downloaded from there. The Galaxy framework and software is open-source which means that is available to everyone and can be run on any Unix-based operating system. Server administration and tool development is supported by an application programming interface (API), software development kit (SDK) and tools for automating set-up and deployment (Goecks, Nekrutenko, and Taylor, 2010). In recent years, researchers have contributed their time, skills and expertise to building this project. These researchers include the members of the Galaxy

users, developers and admins communities. The availability of training materials, workshops and meetings continue to build a strong Galaxy collaboration network around the world (Goecks, Nekrutenko, and Taylor, 2010).

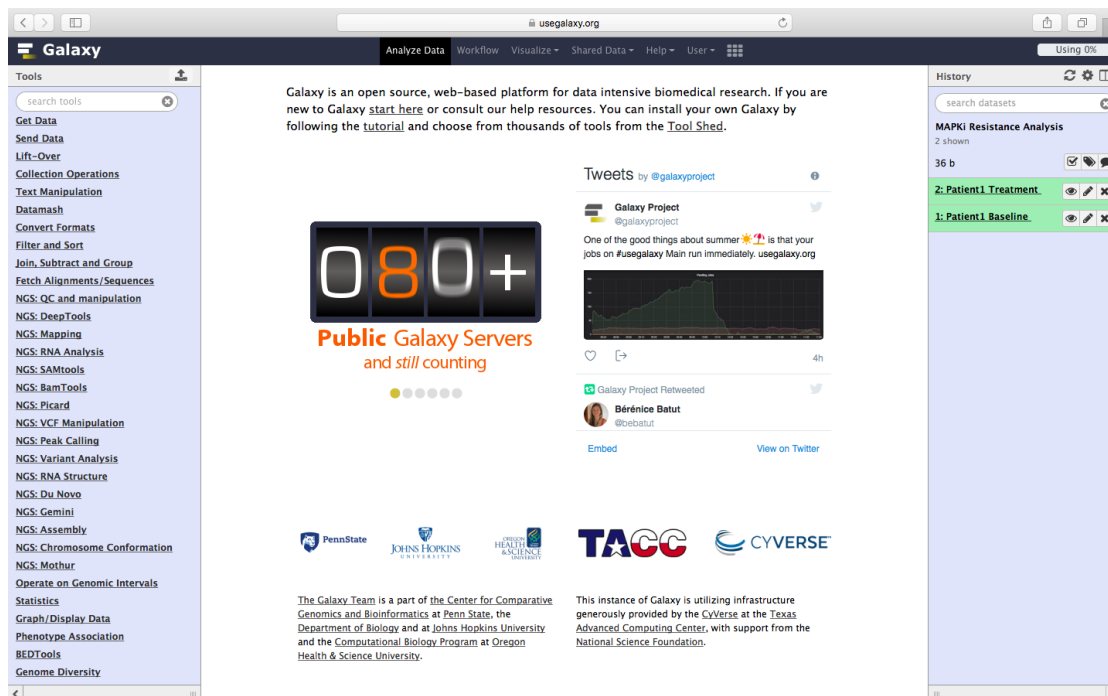


Figure 1.2: Galaxy Project (public) interface consisting of tool search panel, working space area and history.

1.4 Public Databases for Epigenetic Information

To date, epigenetics data are widely available on public resources such as Gene Expression Omnibus (GEO) (Edgar, Domrachev, and Lash, 2002) or ArrayExpress (Brazma et al., 2003) repositories. The number of experiments and their results have been increasing rapidly over the last decade. Current efforts are focused on the extraction and creation of epigenome specific databases. In 2003, the ENCODE Project was launched with the goal of producing a public genome-wide Encyclopedia of DNA Elements (Consortium, 2004). This encyclopedia produces, organizes and also analyses the data, and this project has surpassed the highest expectations of the creators. ENCODE provides not only data but also tools to search and visualize

analysis products in an integrative way. However, the project has been criticized for missing functional elements of genetic and biochemical approaches. In terms of epigenome modifications type there are three types of epigenetic databases include DNA methylation, histone modification and Non-coding RNA mechanism (Holliday, 2006). The DNA methylation databases are important for studying methylomes obtained from different techniques and tissues, pathological conditions, and species. By way of illustration, NGSmethDB is a database for NGS at single cytosine resolution level of DNA methylation data (Hackenberg, Barturen, and Oliver, 2010). Furthermore, the DiseaseMeth database focuses only on human diseases (Lv et al., 2011), while MethylomeDB it is a brain tissue specific database (Xin et al., 2011). Histone repositories are useful for cross-species research in chromatin interactions of genomic DNA, post-translational modifications, and histone modifying enzymes. This approach can be seen in the case of the two freely available databases. Histome database covering around 50 human histone proteins and 150 histone modifying enzymes (Khare et al., 2011) and 4DGenome which instead records multiple chromatin interaction data around 4,433,071 experimentally-derived and 3,605,176 computationally-predicted interactions in 5 organisms.(Teng et al., 2015). Noncoding RNA repositories include target prediction algorithms and experimentally verified miRNA targets. MiRWalk has been developed only for the prediction of possible miRNA binding sites (Dweep et al., 2011). Nevertheless, researchers provide us with miRBase microRNA sequences, targets and gene nomenclature repository (Griffiths Jones et al., 2006). Circular RNAs are specific of they origin from otherwise protein-coding genes and have been shown to be expressed in eukaryotic cells that was the aim for creating circBase containing unique circRNA (Glazar, Papavasileiou, and Rajewsky, 2014). Data mining of genomic regions methylated in cancer have resulted in the creation of cancer limited databases. Customised repositories are useful for cohort studies where irregular methylation patterns that are correlated with various cancers. A leading example of this is The Cancer Genome Atlas (TCGA), which is a publically

funded project that aims to accelerate understanding of cancer genetics by cataloging and storing of cancer genomic profiles (Tomczak, 2015). Other comprehensive databases available for DNA methylation and cancer include MethyCancer (He et al., 2007) and a database of 167 epigenetic modifiers correlated with cancer targets (Nanda, Kumar, and Raghava, 2016). These computational strategies and resources offer new opportunities for greater understanding of epigenome regulation, molecular organization, development and disease (Gautam et al., 2018).

Chapter 2

Computational Methods for Epigenetics

2.1 Epigenome-Wide DNA Methylation Profiling

Epigenome-wide association studies (EWAS) analyse genome-wide distribution of epigenetic marks in cohorts of different individuals in order to find associations between epigenetic variation and phenotype (Rakyan et al., 2011). One of the epigenetic modifications widely studied is DNA methylation. In humans, DNA methylation occurs by attaching a methyl group to the cytosine residue and it has been suggested that this modification results in as a suppression of gene expression (Bird, 1986). With regards to DNA methylation analysis there are a range of technologies such as, polymerase chain reaction (PCR) or pyrosequencing which are dedicated to studying small groups of methylation sites across number of samples (Kristensen and Hansen, 2009). New approaches, including bisulphite sequencing (RRBS) (Carr et al., 2007) and sequencing by synthesis (MethylC-Seq) (Urich et al., 2015), have allowed researchers to study DNA methylation on the global scale. Results from RRBS and MethylC-seq are comparable with one another (Urich et al., 2015). A potentially significant problem with these methods is its level of complexity and cost escalating with population sizes, meaning that currently running this type of analysis can be prohibitively costly (Kristensen and Hansen, 2009). By contrast, Illumina Methylation Assay (Illumina, 2018) provides high accuracy

with low input DNA requirements on budget. Bisulfite conversion, array processing, and a wide study of methylation changes with other platform compatibility, (e.g. gene expression, microRNA profiling) make Illumina Methylation Assay one of the most comprehensive solution on the market (Sandoval et al., 2011). However, a Illumina Genome Studio software license is required, which may not be suitable for everyone. Moreover, as the company offers only basic preprocessing and analysis options there is growing interest to create freely available software to perform quality control, normalization and detection of differentially methylated regions (Marabita et al., 2013a). Existing tools and pipelines also require high performance computational hardware, programming knowledge and experience to run the analysis. This is why the aim of my project was to establish and implement this methods into user-friendly environment.

There are two basic approaches currently being adopted in research for reading the 450k data. First focuses on efficiency of bisulfite conversion and the second on the overall experiment (e.g. hybridization, extension). Illumina Methylation Assay 450k is a powerful utilisation in terms of reagent costs, time of labour, high accuracy, low input DNA requirements and price. It determines quantitative array-based methylation high resolution measurements at the single-CpG-site level of over 450 thousand loci (Pidsley et al., 2013). Epigenome-wide methylation analysis capabilities make this assay suitable for broad investigation of methylation changes in normal and diseased cells (Rakyan et al., 2011). Infinium technology uses two different bead types to detect changes in DNA methylation levels. In Figure 2.1 one can see M - methylated and U - unmethylated bead types. Depending on the probe design, the bead signals are reported in different colours -green or red (Illumina, 2018).

As such it has become one of the most comprehensive solutions on the market (Marabita et al., 2013a). However, due to the nature of the design following two different chemical assays, analysis can be too complex. As mentioned previously, the Illumina Genome Studio software license is required

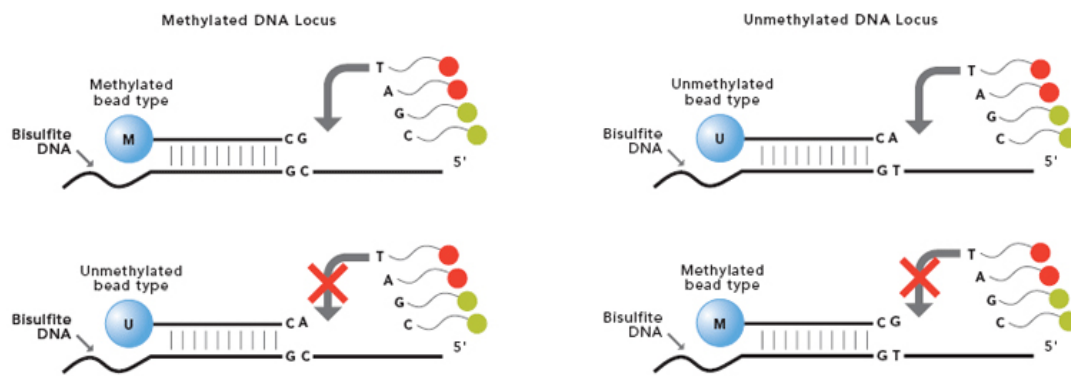


Figure 2.1: The Infinium Methylation Assay Technology (Illumina, 2018)

before use and the company offers only basic preprocessing and analysis strategies (Marabita et al., 2013a). 450k probes first needs to be filtered and normalised. Filtering out any probes that can generate artifactual data and removing any source of variation to avoid any data deviations. After correct preprocessing of the data (i.e. filtering out problematic probes and normalizing the data), downstream analysis can be performed. Detection of DMPs (differentially methylated positions) and DMRs (differentially methylated regions) is the most attractive currently available technique in the framework of large biomarker discovery studies, and has resulted in a growing interest to create freely available software to perform quality control, normalization and differentially methylated regions detection (Marabita et al., 2013a).

2.1.1 Pre-processing

No single technique excels in all aspects of DNA methylation analysis. Sample number and associated characteristics, and the method applied will generate different types and number of results. Therefore, it is necessary to understand, identify, and exclude low quality scores. Data preprocessing may be divided into normalization and filtering step. Different methods have been proposed to perform normalization. The Infinium I probe design includes two bead types per SNP locus and ends in the queried SNP base (Nakabayashi, 2017). Infinium II probe design requires just one bead type per locus and

ends at the base proceeding of SNP (Nakabayashi, 2017). Both of designs allow unlimited access for SNP interrogation and minimize the differences so ranking of potential differentially methylated loci is more accurate (Illumina, 2018). Background and inter-array correction equalizes the background signal between Type I and Type II probes and helps to achieve good quality data especially when subtle methylation differences need to be detected by statistical tests between large numbers of cases and controls. Filtering otherwise provides a basis for detection and removal of SNP variation (Daca-Roszak et al., 2015), low p-values and cross-reactive events (Marabita et al., 2013b). This helps to correct for possible bias arising from within and between array variation which normalization did not detect (Aryee et al., 2014). Computational methods can be applied to sequencing datasets via command line tools or R based trimming toolkits, such as Bioconductor package Minfi and its Quality Control functions (Aryee et al., 2014). Here function return plots with the log median intensity in both the methylated (M) and unmethylated (U) channels. When plotting these two medians against each other the good quality samples cluster together, while failed samples tend to separate and have lower median intensities (Aryee et al., 2014). In addition Bioconductor offers us tailored function to remove probes that contain either a SNP at the CpG interrogation or at the single nucleotide extension (Aryee et al., 2014).

2.1.2 Mapping and Graphical Visualization

Generally, downstream data analysis provides three types of information: statistical significance between groups, absolute differences (β , M value) and differentially methylated areas (Aryee et al., 2014). Beta value ranges from 0 to 1 indicating the methylation level at that site for the population of cells analysed (0 - unmethylated, 1 - methylated) (Cazaly et al., 2016). Unfortunately at very high or low values there is a risk of unequal variance at β values record. To avoid any heteroskedasticity β values can be transformed into the logit - M value (Cazaly et al., 2016). Differentially methylated areas involves adjacent positions or regions that have different methylation patterns

between samples (Aryee et al., 2014). The first step in high-throughput sequencing analysis is the mapping of the generated reads to a genome reference sequence. To accomplish this task, numerous software tools have been developed. Alignment is the process where data are mapped to the genome. Depending on the genome, not all methylation loci may have a genomic position. Currently, popular software alignment tools use UCSC genome browser (<http://genome.ucsc.edu>) (Kent et al., 2002) to optimally match the reference genome with data. The UCSC Genome Browser site provides reference sequences and draft assemblies for various genomes. Graphical viewers have been developed to support fast interactive performance. Open-source, web-based tool suite for genomic visualization, examination and curation of the custom data. Infinium methylation arrays can be associated with a genomic location using Minfi package (Aryee et al., 2014). Data mapped to genome can be generated or converted in normalized BedGraph format e.g. using Bioconductor rtracklayer package. Rtracklayer is a framework for interacting with genome browsers and manipulating annotation tracks (Lawrence, Gentleman, and Carey, 2009). BedGraph format allows display of continuous-valued data in track format which can be visualized in genome browsers or genomics viewers (Kent et al., 2002). Graphical representation of data give better understanding of generated reads and their possible implication.

2.1.3 DMR and DMP Identification

The next step of methylation profiling is identification of differentially methylated loci with respect to case or control design. In simple comparisons between such pairs of samples (e.g. treatment and control) Fisher's Exact Test or Hidden Markov Models (HMMs) are adequate to compare one test and one control sample at a time. They are implemented in Bioconductor packages such as methylKit (Akalin A, 2012), RnBeads (Assenov et al., 2014) and VanillaICE (Scharpf RB, 2008). These studies would have been more useful if they had focused on replicates variation. Replication between a discovery and a validation data set improve the measurement of variation and

hence increases the precision of gene expression measurements and allows smaller changes to be detected. Genomic regions (DMR) that are differentially methylated between replicates can be tracked using regression bump hunting framework (Jaffe et al., 2012). The algorithm first implements a t-statistic at each methylated loci location, with optional smoothing, then grouping probes into clusters with a maximum location gap and a cutoff size to refer the lowest possible value of genomic profile hunted (Jaffe et al., 2012). Differentially methylated positions (DMP) detection can be accomplished by using linear regression model from limma package (Ritchie ME, 2015). Limma was initially developed for the analysis of microarray expression analysis but currently it is also used for methylation data (Ritchie ME, 2015). In order to reveal the meaning of methylation changes results can be functionally annotated or correlated with the corresponding expression data.

2.1.4 Peak Detection

Peak Detection is a computational method added and used in methylation analysis to identify areas where a protein (e.g., modified histones or transcription factors) interacts with DNA (Pepke, Wold, and Mortazavi, 2009). Transcription factors enriched area are called transcription factor binding sites (TFBS), while for histone modifications enriched area are referred to as histone modification peak (Pepke, Wold, and Mortazavi, 2009). In DNA methylation experiments, an area of interest is the methylated enriched region (Klose and Bird, 2006). Software packages for peak detection follow these basic components: first, call peaks along individual chromosome than combine individual signals to apply post-call filtering and statistical tests. Significance ranking of called peaks estimate regulatory sites or whole different distribution patterns. There are two preferred approaches for mapping puncture peaks e.g. sites of transcription-factor binding or methylation levels. These are MACS - Model-Based Analysis (Zhang et al., 2008) and PeakSeq systematic scoring (Rozowsky et al., 2009). Broad and narrow peaks e.g. open chromatin regions can be detected using F-Seq (Boyle, 2008) or WaveSeq

(Mitra A, 2012). The key problem with peak calling algorithms and tools, is that customized settings and parameters can affect the number of peaks and potential can confuse the user. The power and abundance of counting-based measurements create new challenges and features for future analysis tools.

2.2 Galaxy Tools Development

Galaxy is an open-source project. Everyone can contribute to its development as it is not focused on any single software but rather enables integration of different technologies (Afgan et al., 2018). In addition, Galaxy allow researchers to share code and work in straightforward manner without being concerned with what programming language or environment was used (Afgan et al., 2018). The XML file is a link between the tool and Galaxy which describes to Galaxy how the underlying software works i.e., how to invoke the tool, what options to pass and what the tool will produce as output. Dependency and libraries required for installation can be added via Conda (Anaconda, 2017). Conda quickly installs, runs and updates packages and software requirements for any language — Python, R, Ruby, Lua, Scala, Java, JavaScript, C/ C++, FORTRAN (Anaconda, 2017). Functional testing is a recipe quality control process to present developers and users with tools that can be run across different systems and architectures. Tools are tested by feeding them with example datasets and settings and then examining the results. Every Galaxy dataset is associated with a datatype which can be added to the config file if the latter it is not already there (Afgan et al., 2018). Appropriate types of the data have great importance for the analysis e.g. SAM/BAM, BED, GFF/GTF, WIG, bigWig, bigBed, bedGraph, and VCF offer specific visualization and visual analysis environment options available via Trackster (Afgan et al., 2018). Galaxy's aim is the integration of datasets easily without downloading or sending it to a remote server. Following this idea developers can make tools accessible to a broad audience by sharing and publishing them on the main server accessible to everyone (Afgan et al., 2018).

The tools and tool suites provide ready-to-use combinations of bioinformatics methods that are wrapped into Galaxy ToolShed (Poterlowicz, 2018a). The ToolShed allows Galaxy administrators to install thousands of freely available Galaxy utilities into their instances. The main ToolShed serves as a digital distribution platform, sharing tool updates and versions. This solution simplifies management of tools for both developers and administrators (Afgan et al., 2018). Tools can be developed separately in respect to researchers needs. The goal of creating groups of tools and build them into suites is to expand both their quantity and quality (Goecks, Nekrutenko, and Taylor, 2010). The number of Galaxy tools contributed by the community has increased as a result of growing interest in multiple areas of research. In Galaxy, users can find simple statistics or even text manipulation tools so there is no need to download your data and edit them manually, but the majority of tools on the instance are for analysis of Next-generation sequencing (NGS) genomic datasets. In addition to newly available tools, Galaxy recognizes and can process data types from current DNA sequencers (Goecks, Nekrutenko, and Taylor, 2010). Galaxy implements a range of attributes to simplify the analysis of big data, including workflows and collections. Users are able to analyze their data in interactive and reproducible ways. Workflow's trial-and-error approach allows them to use individual tools in following order and connection. Pipelines can be generated from history or downloaded from external resources (e.g. (Manchester and Southampton, 2018)). A convenient workflow editor shown on Figure 2.2 is also available to build workflows step-by-step or to edit existing ones. Galaxy gives user flexibility by providing named-tags and labels for tools and datasets used in the analysis. Once developed, workflows behave complex tools, and they can be obtained and executed from Galaxy's main analysis interface (Goecks, Nekrutenko, and Taylor, 2010).

However, local galaxy instance requires a few additional things to run: a virtual environment, configuration files, and dependent Python modules (Afgan et al., 2018). However, starting the server requires informatics expertise

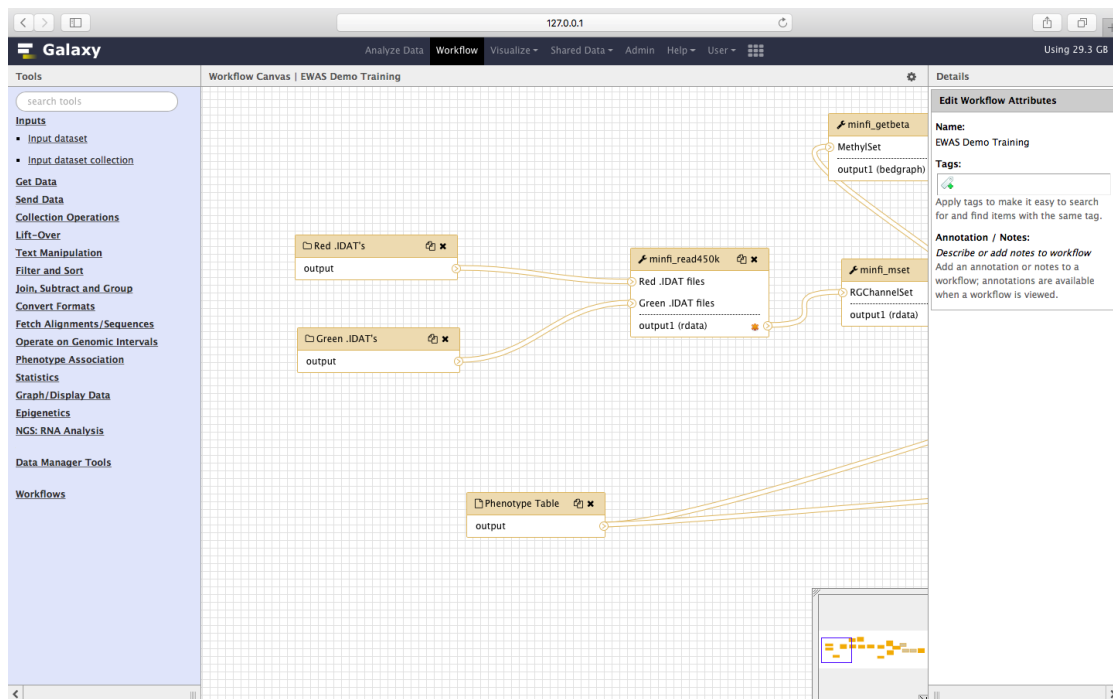


Figure 2.2: Galaxy's graphical workflow editor, show part of a sample workflow.

to set up infrastructure and managing the instance. Firstly users should install Galaxy, tools and all the necessary dependencies on their local machine. Then administration work starts when maintaining the server requests not only user handling but also controlling the usage, data back-up and software updates. Galaxy provides multiple benefits from local installation. Users can customize Galaxy itself as can be seen on Figure 2.3. This shows a "kp-bioteam" instance being run by Dr K Poterlowicz team. In addition, the admin settings section allows installation of the tools not to be published onto the main server or to develop new tools that integrate with particular research question. Scalability optimize runs, handle more users, run more jobs even on large datasets. Disadvantages on the informatics side are fully addressed by various advantages of this solution (Afgan et al., 2018).

Planemo are command-line utilities assist in developing Galaxy and Common Workflow Language tools. The key aspects of Planemo virtual appliance are availability to local development environments (e.g. if Planemo has been installed with brew or pip) and the second is for developers using a dedicated Planemo virtual appliance (available as OVA, Docker, Vagrant, etc.)

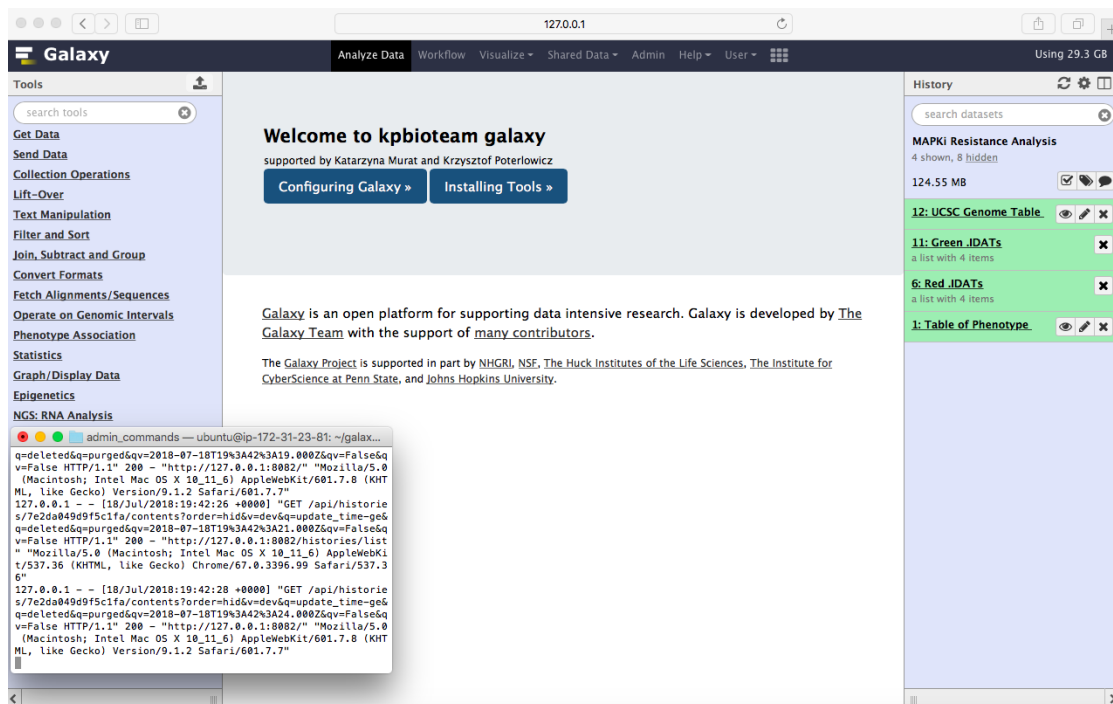


Figure 2.3: Local Galaxy Instance running via command line with custom interface consisting predefined tools sections, tailored working space area and test history.

(galaxyproject, 2014). Planemo can help user to develop tools and Conda packages in unison, publish tools to the Galaxy Tool Shed and also support docker and containers (galaxyproject, 2014). Docker, a virtual appliance is an open-source technology that performs operating-system-level virtualization also known as containerisation (Developers, 2017). It enables developers to easily pack, ship, and run any application as a lightweight, portable, self-sufficient container, which can run virtually everywhere. Containers are created from runtime instance of an image which specify their exact contents. Images are often created by combining and modifying standard images downloaded from repositories (Developers, 2017). In addition, Docker (Developers, 2017) can get more applications running on the same time and more hardware than other computational solutions. Simplified containered applications are ready-to-run by developers and it makes managing and deploying much easier. Nowadays, almost all IT and cloud companies have adopted Docker which means that the community is still growing and it is really valuable to use it (Developers, 2017).

Chapter 3

Development of Galaxy Tools

3.1 EWAS - Galaxy Tool Suite

An EWAS suite (table 3.1) has been developed as a part of my project to provide users with an enhanced understanding of the Infinium Methylation Assay analysis tool. The tool suite includes methods for preprocessing with stratified quantile normalization `minfi_ppquantile` or extended implementation of functional normalization `minfi_ppfun` with unwanted variation removal, sample specific quality assessment `minfi_qc` and differentially methylated regions `minfi_dmr` and position detection `minfi_dmp`. All scripts were wrapped into a web based platform - Galaxy, as a user-friendly interface with tools and ready to run workflows. Which is a solution for non-programmer scientists allowing them to analyze their data and share their experience with others (Poterlowicz, 2018a). Configuration files are publicly published on our lab group's GitHub repository with scripts and dependency settings also available to download and install via the Galaxy test toolshed (Poterlowicz, 2018a). My suite was created and tested using a Planemo workspace with a default configuration and shed tool setup available via Docker (operating-system-level virtualization) (Poterlowicz, 2018b).

The workflow combines 7 main steps Figure 3.1, starting with raw intensity data loading (`/idat`) and then preprocessing and optional normalization of the data. The next quality control step performs an additional sample check to remove low-quality data, which normalization cannot detect. The workflow gives the user the opportunity to perform any of these preparation and

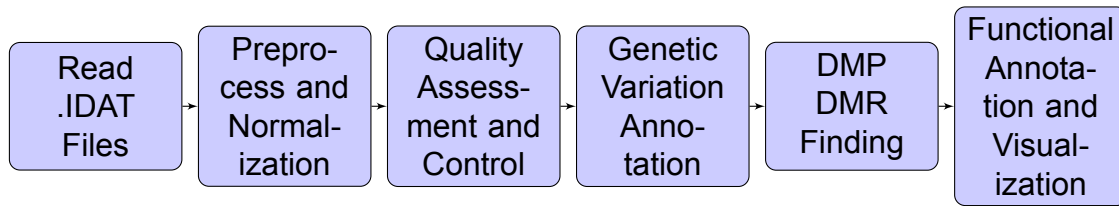


Figure 3.1: Simplified workflow for analysing epigenetics data

data cleaning steps, including the highly recommended genetic variation annotation step that results in single nucleotide polymorphism finding and removal. Finally, the dataset generated through all of these steps can be used to hunt (find) differentially methylated positions (DMP) and regions (DMR) with respect to a phenotype covariate. Functional annotation of these data generates clinically meaningful information about methylation changes with graphical representation of these genes and functions. All the tools, single preparation and analysis steps are shown in Figure 3.2 and explained in detail below.

Data Loading: IDAT files are the combination of raw green and red colour arrays containing the summarized bead information generated by the Illumina 450k scanner (Marabita et al., 2013a). Illumina’s GenomeStudio solution converts the data into plain-text ASCII files losing a large amount of information during this process (Marabita et al., 2013a). To prevent this kind of data loss we developed an R based tool `minfi_read450k` which is the combination of `illuminaio readIDAT` and `minfi RGChannelSet` function. We decided to use this functions to firstly load intensity information for each two color micro array and then build up an `RGChannelSet` class based on them. **Preprocessing and Normalization:** `RGChannelSet` represents two color data with a green and a red channel and can be converted into methylated and unmethylated signals assigned to `MethylSet` or into Beta values build in `RatioSet` (Aryee et al., 2014). Users can convert from `RGChannelSet` into `MethylSet` using the `minfi_mset` tool or compute Beta values using `minfi_rset` tool, no normalization performed. However, these two classes can also be preprocessed and normalized with two methods recommended by Illumina (Aryee et al., 2014). `Minfi_ppquantile` implements stratified quantile normalization preprocessing

Table 3.1: Summary of the EWAS suite tools inputs and outputs

Tool	Input	Output	Description
minfi_read450k	IDAT	RGChannelSet	read the .IDAT files
minfi_mset	RGChannelSet	MethylSet	convert the Red/Green .IDAT's for an Illumina methylation array
minfi_qc	MethylSet / GenomicMethylSet	DataFrame	quality assessment
minfi_rset	MethylSet/GenomicRatioSet	RatioSet	converting methylation data from methylation and unmethylation channels, to ratios (Beta and M-values)
minfi_ppfun	RGChannelSet	GenomicRatioSet	functional normalization preprocessing
minfi_ppquantile	RGChannelSet/GenomicMethylSet	GenomicRatioSet	stratified quantile normalization
minfi_maptogenome	MethylSet/RGChannelSet/RatioSet	GenomicRatioSet	add genomic coordinates to each probe together with some additional annotation information
minfi_geo	GEO accession	GenomicRatioSet	download data from GEO database
minfi_getbeta	MethylSet/RatioSet/GenomicRatioSet	DataFrame	return Beta value
minfi_getCN	MethylSet/RatioSet/GenomicRatioSet	DataFrame	return coordinating node
minfi_getM	MethylSet/RatioSet/GenomicRatioSet	DataFrame	return the Fisher information corresponding to a model and a design
minfi_pheno	RatioSet/GenomicRatioSet	DataFrame	extract phenotype data
minfi_getanno	GenomicRatioSet	DataFrame	access provided annotation
minfi_getsnp	GenomicRatioSet	DataFrame	return SNP information of the probes
minfi_dropshp	GenomicRatioSet	GenomicRatioSet	drop the probes that contain either a SNP at the metylated loci interrogation or at the single nucleotide extension
minfi_dmp	MethylSet/GenomicRatioSet	DataFrame	return differentially methylated positions
minfi_dmr	GenomicRatioSet	DataFrame	return differentially methylated regions

and it is supported for small changes like in one-type samples e.g. blood dataset. The conceptual understanding is that it is a transformation of the array x based on function $F^{-1}Gx$ where Gx is an estimated distribution function and F^{-1} is the inverse of an estimated distribution function (Bolstad et al., 2003). It has the result in normalized distributions identical for all the arrays (Bolstad et al., 2003). For quantile normalization Gx is the empirical distribution of array x and F is the empirical distribution for the averaged quantiles across arrays (Bolstad et al., 2003). To sum up, quantile normalization is a technique for making two arrays distributions identical in statistical properties. In contrast, `minfi_ppfun` is aimed at global biological differences such as healthy and occurred datasets with different tissues and cell types (Fortin et al., 2014). It is called the between-array normalization method and removes only variation explained by a set of covariates while quantile normalization forces samples to be the same, which removes all variation across the data (Fortin et al., 2014). Both of these methods return `GenomicRatioSet` class aimed at most of analyses performed on EWAS data using `Minfi` package (Aryee et al., 2014).

Quality Assessment and Control: Data quality assurance is an important step in Infinium Methylation Assay analysis. The assessment can be run on methylated signals generated by preprocessing analysis step. `Minfi_qc` tool outputs plot the log median intensity in both the methylated (M) and unmethylated (U) channels. When plotting these two medians against each other the good samples cluster together, while failed samples tend to separate and have lower median intensities as we can see on Figure 4.1 all provided samples are remarkable.

Mapping: Mapping is the process where a sequenced read is compared to a reference based on its nucleotide sequence similarity. `Minfi_maptogenome` tool accept both `Methyl-` and `RatioSet` than align to the genome using an annotation package and output `GenomicRatioSet` or `GenomicMethylSet`. However, depending on the genome, not all methylation loci may have a genomic position.

Annotating probes affected by genetic variation: Incomplete annotation of genetic variations

such as single nucleotide polymorphism (SNP) may affect DNA measurements and interfere results from downstream analysis. `Minfi_getsnp` return a data frame containing the SNP information of unwanted probes to be removed by `minfi_dropsnp` tool (Aryee et al., 2014). **DMPs and DMRs Identification:** The main goal of the EWAS suite is to simplify the way differentially methylated loci sites are detected. The EWAS suite contains `minfi_dmp` tool detecting differentially methylated positions (DMPs) with respect to a phenotype covariate, and more complex `minfi_dmr` solution for finding differentially methylated regions (DMRs) (Aryee et al., 2014). Genomic regions that are differentially methylated between two conditions can be tracked using a bump hunting algorithm. The algorithm first implements a t-statistic at each methylated loci location, with optional smoothing, then groups probes into clusters with a maximum location gap and a cutoff size (Aryee et al., 2014). **Functional Annotation and Visualization:** In addition to downstream analysis users can access annotation provided via Illumina (`minfi_getanno`) (Aryee et al., 2014) or perform additional functional annotation using the Gene Ontology (GO) tool (`clusterprofiler_go`). The Gene Ontology (GO) provides a very detailed representation of functional relationships between biological processes, molecular function and cellular components across data (Consortium, 2004). Once a specific regions has been chosen, `clusterprofiler_go` visualize enrichment result see Figure 4.2. Many researchers use pathway analysis to characterise the function of the gene that demonstrate the potential for Galaxy to be a bioinformatics solution for wide multi-omics research. **Training:** I have also provided training sessions and interactive tours for user self-learning. The training materials are freely accessible at the Galaxy project Github repository (<http://galaxyproject.github.io/training-material>). Such trainings and tours guide users through an entire analysis. Following steps and notes helps users to explore and better understand the concept. Slides, a hands-on instruction describes the analysis workflow, all necessary input files are ready-to-use via Zenodo, a Galaxy Interactive Tour, and a tailor-made Galaxy Docker image for the corresponding data analysis.

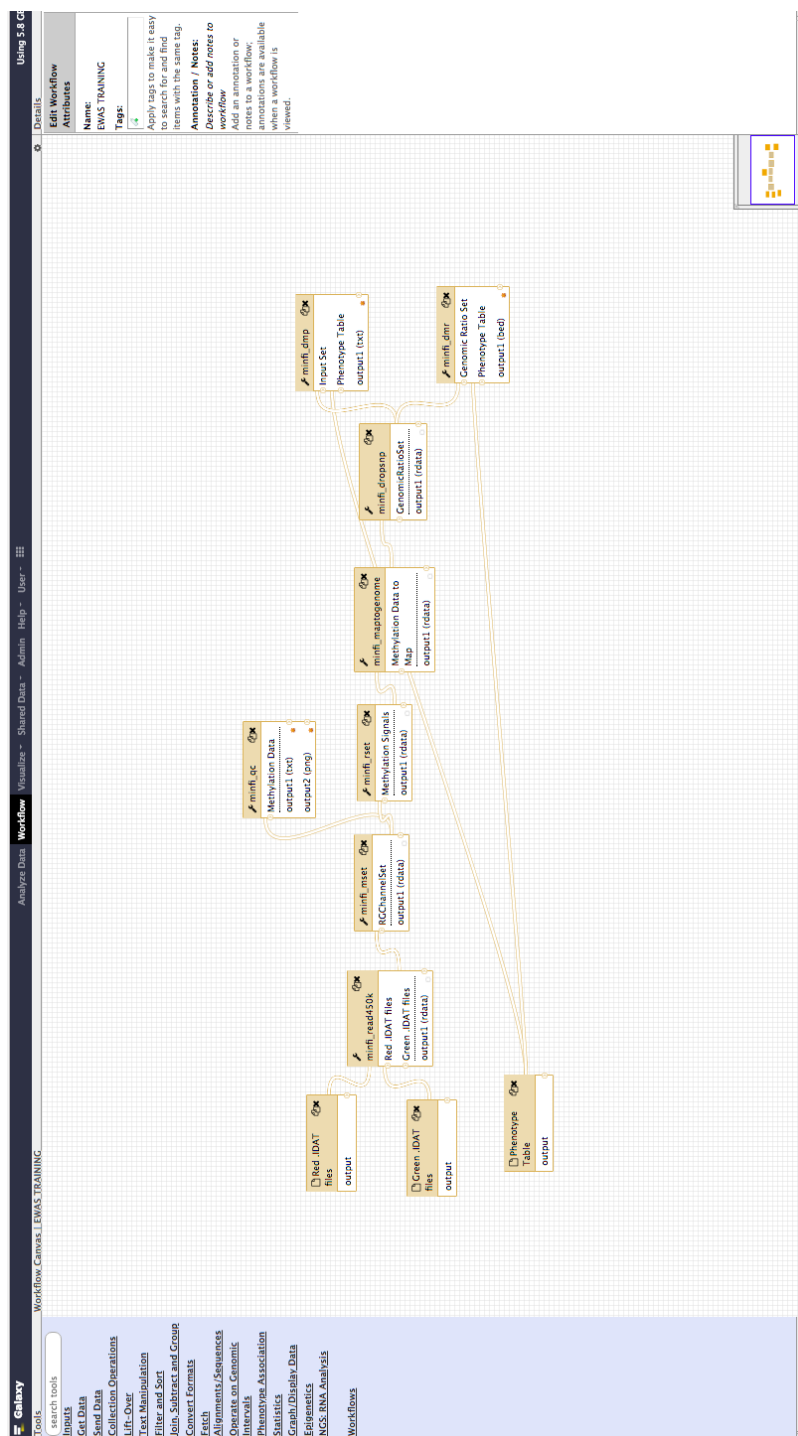


Figure 3.2: Screenshot from the Galaxy Workflow Editor, showing EWAS example workflow discussed in the Analyses section.

Chapter 4

Application of the Tools

4.1 Enhancement for the Computational Methods

With the rapidly increasing volume of epigenetics data available, a computer-based analysis of heritable changes in gene expression becomes more and more feasible. Many computational epigenetics studies have focused on generation of data and increasing diversity of methods and techniques on data population scale and tools to mine them. The generalization of much of the published research on this issue is however, problematic. It is true that executing and aligning for multi-omics studies are important steps. Risk evaluation, disease management and novel therapeutics development are all challenging researchers to find novel bioinformatic frameworks and approaches. Recent work has revealed that this challenge is now being targeted (Holliday, 2006). The main difference to existing solutions is that they require computing knowledge and experience. In this regard I provide a well established user friendly tool suite available via Galaxy platform 'EWAS-Galaxy'. It combines theories, models and methods required to run complex biological and medical epigenetics analysis. In addition, it provides a set of tools, each one being available as a BioConda package as well as a Docker container (Poterlowicz, 2018b). Based on the Galaxy Docker project, my web server offers a comprehensive and freely accessible epigenetics workspace. Source code for the Galaxy is open and supported by the developer community, which means that my tools are being tested and constantly improved. They can be deployed on every standard operating system (Linux, Windows, OSX), but at

the same time they can be installed and run on high-performance- or cloud-computing infrastructure (Poterlowicz, 2018a). The role of computational epigenetics is the development and application of bioinformatics methods to study the epigenome.

4.2 Clinical Relevance Validation

A Big interest in skin cancer biomarker identification led to validation of the differentially methylated regions analysis. Illumina 450K Methylation data were obtained for sensitivity of melanoma biopsies pre and post MAPKi treatment (Hugo et al., 2015). The data has been download from Gene Expression Omnibus (GEO) with accession number GSE65183. The Gene Expression Omnibus is an international public database repository which distributes broadly understood genomic data sets (Hugo et al., 2015). Methylation profiling by genome tiling array in melanoma can help us understand how non-genomic and immune changes can have an impact on treatment efficiency and disease progression. Raw image IDAT files were loaded into the Galaxy environment using Data Libraries. EWAS workflow was run on Red and Green dataset collections of patient-matched melanoma tumours biopsied before therapy and during disease progression, pre-defined phenotype tables with sensitivity information and up to date genome tables (UCSC Main on Human hg19 Methyl450) (Poterlowicz, 2018b) were used as inputs and default settings. This workflow generated differentially methylated regions and positions and also studied the functional aspects behind hypo- and hyper - methylated genes. To detect poorly performing samples I ran quality diagnostics with minfi_qc tool. Provided samples passed quality control test, can be seen (on figure 4.1) that they clustered together with higher median intensities confirming their good quality (Aryee et al., 2014).

Differentially methylated loci were identified using single probe analysis implemented by **minfi_dmp** tool with the following parameters: phenotype set

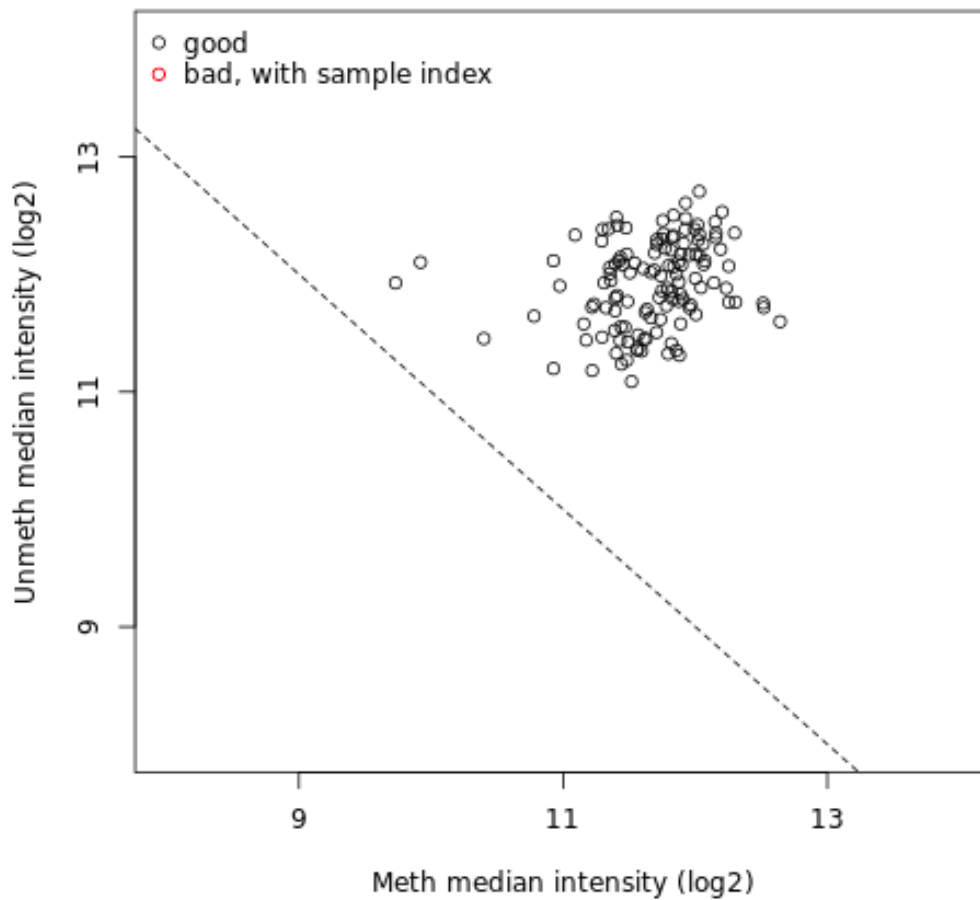


Figure 4.1: Quality Control Plot representation of melanoma pre and post MAPKi treatment samples.

as **categorical** and qCutoff size set to **1**. The bump hunting algorithm was applied into **minfi_dmr** tool to identify differentially methylated regions (DMRs) with maximum location gap parameter set to **250**, genomic profile above the cutoff equal to **0.1**, number of resamples set to **0**, null method set to **permutation** and verbose equal **FALSE** which means that no additional progress information will be printed. Differentially Methylated Regions and Positions revealed the need for further investigation of tissue diversity in response to environmental changes (Bock and Lengauer, 2008). Nearest transcription start sites (TSS) and enhancer elements annotations were found in the hypermethylated gene set listed as follows: PITX1, SFRP2, MSX1, MIR21, AXIN2, GREM1, WT1, CBX2, HCK, GTSE1, SNCG, PDPN, PDGFRA, NAF1, FGF5,

FOXE1, THBS1, DLK1 and HOX gene family.

Functional annotation with GO is a schema to understand how the annotations are assigned to the genes (Ashburner et al., 2000). These are enrichment GO categories after controlling for FDR control Figure 4.2. The most significance to the gene output was the pattern specification process (GO:0007389), skeletal system development (GO:0001501) and regionalization (GO:0003002) meaning that melanoma MAPKi resistance could be related to the cells developmental process within specific environments.

4.3 Discovery and Validation of a Predicted SKCM Epigenetics Variants

Epigenetic aberrations that involve DNA modifications give researchers the possibility to identify novel non-genetic factors responsible for complex human phenotypes such as height, weight, and disease. To identify methylation changes researchers need to perform complicated and time consuming computational analysis (Bock and Lengauer, 2008). Here, the EWAS suite becomes a solution for this inconvenience and provides a simplified downstream analysis including preprocessing, quality evaluation and differentially methylated CpG site detection in one complex set of tools developed and published under the Galaxy platform. I also show how my initial implementation of EWAS tools suite combination, Figure 3.2, can provide additional insights into e.g. melanoma therapeutic resistance. Workflow published on the kpbioteam docker instance allow users to repeat the analysis performed on melanoma data with their own examples and feedback improvements. Diagnostic biomarkers currently used to assist in the diagnosis of melanoma were founded in chosen dataset. However, the study fails to consider the differing categories of treatment. Recent cases reported by (Hugo et al., 2015) showed that gene and signature based transcriptomic alterations in acquired MAPKi resistant melanoma were highly recurrent. This can help to explain clinical

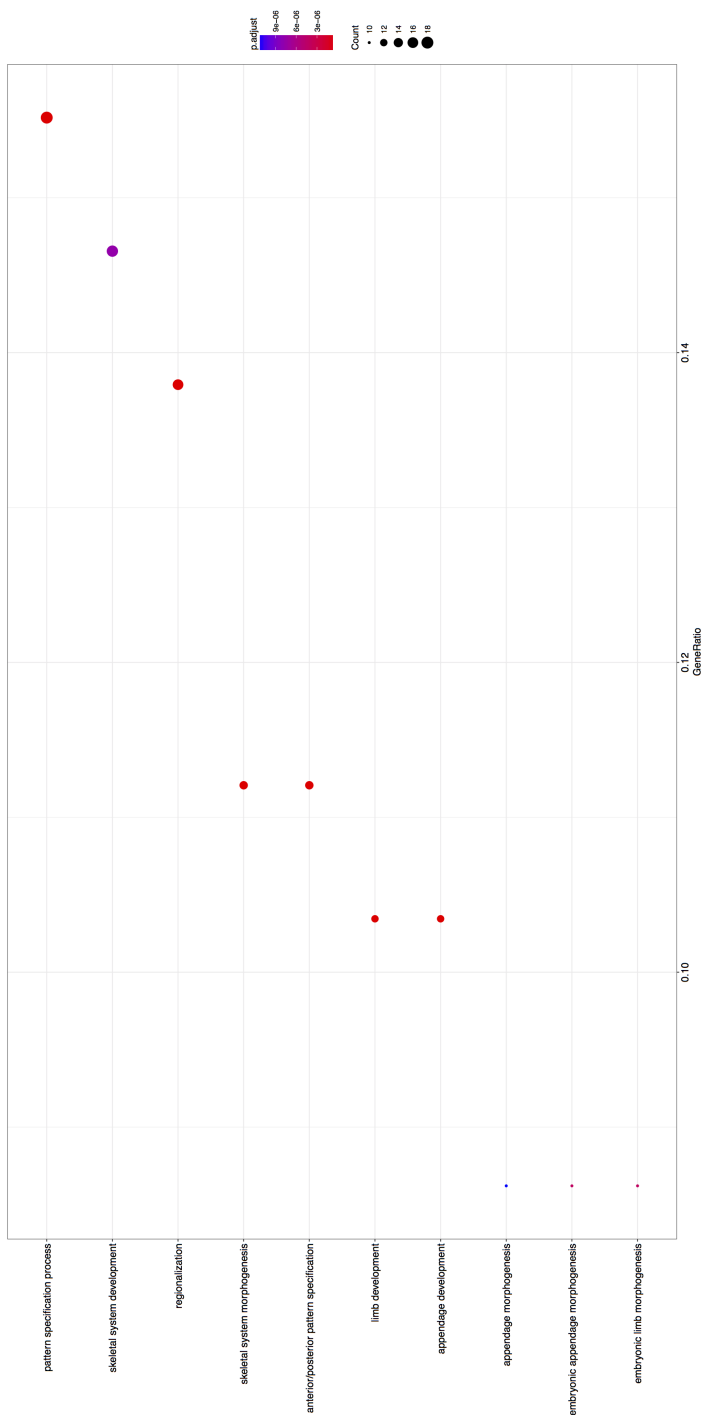


Figure 4.2: Functional Annotation of DMR's found in melanoma biopsies pre and post MAPKi treatment.

relapse comprehensively with new genetic and epigenetic mechanism findings (Gilbertson et al., 2006). Initial observations suggest that there may be a link between MAPKi resistance and DNA methylation changes itself. Here, we showed that specific genes and pathways subject to differential regulation in resistant tumour cells. I highlighted a group of hypermethylated genes already connected to cancer which a lack of MAPKi research. This data demonstrates that PDGFR, which is suggested to be responsible for RAS/MAPK pathway activation can truly regulate the MAPKi mechanism in non responsive tumours, but its altered methylation regulation requires additional studies (Hugo et al., 2015). Hypermethylation can be associated with expression down-regulation. The PITX1 protein is a member of the bicoid-related homeobox transcription factors and was founded as contributor to the progression of human cutaneous malignant melanoma (Osaki et al., 2013). Previous published studies are limited to local surveys and serial biopsies. Thus, stimulus of innate or acquired MAPKi resistance may converge on epigenetics. I also presented that homeodomain transcription factor MSX1 and CBX2 polycomb protein are likely to be treatment resistance factors, reported as downregulated and inactivated in melanoma tumours (Clermont et al., 2014). Additional analysis performed on MAPKi treatment sensitivity data reveal new potential directions for therapeutic approaches.

Appendix A

Galaxy EWAS Training

450K Array Data Upload

The first step of EWAS - Galaxy data analysis is raw methylation data loading (intensity information files for each two color micro array)

1. Create a new history for this tutorial and give it a proper name
2. Import the following IDAT files from Zenodo ([link](#))
3. Run minfi_read450k with the following parameters for input files A.1
red files: GSM1588704_8795207135_R01C02_Red.idat pre-treatment,
 GSM1588705_8795207119_R05C02_Red.idat pre-treatment,
 GSM1588706_8795207135_R02C02_Red.idat BRAFi treatment resistant,
 GSM1588707_8795207119_R06C02_Red.idat BRAFi treatment resistant
green files: GSM1588704_8795207135_R01C02_Grn.idat pre-treatment,
 GSM1588705_8795207119_R05C02_Grn.idat pre-treatment,
 GSM1588706_8795207135_R02C02_Grn.idat BRAFi treatment resistant,
 GSM1588707_8795207119_R06C02_Grn.idat BRAFi treatment resistant
4. Inspect generated set of data

Preprocessing and Quality Assessment

Preprocessing and data quality assurance is an important step in Infinium Methylation Assay analysis (Aryee et al., 2014). RGChannelSet represents two color data with a green and a red channel and can be converted into methylated and unmethylated signals assigned to MethylSet or into Beta values build in RatioSet. User can convert from RGChannelSet into MethylSet using the minfi_mset or compute Beta values using minfi_set. The minfi_qc tool extracts and plots the quality control data frame with two columns mMed

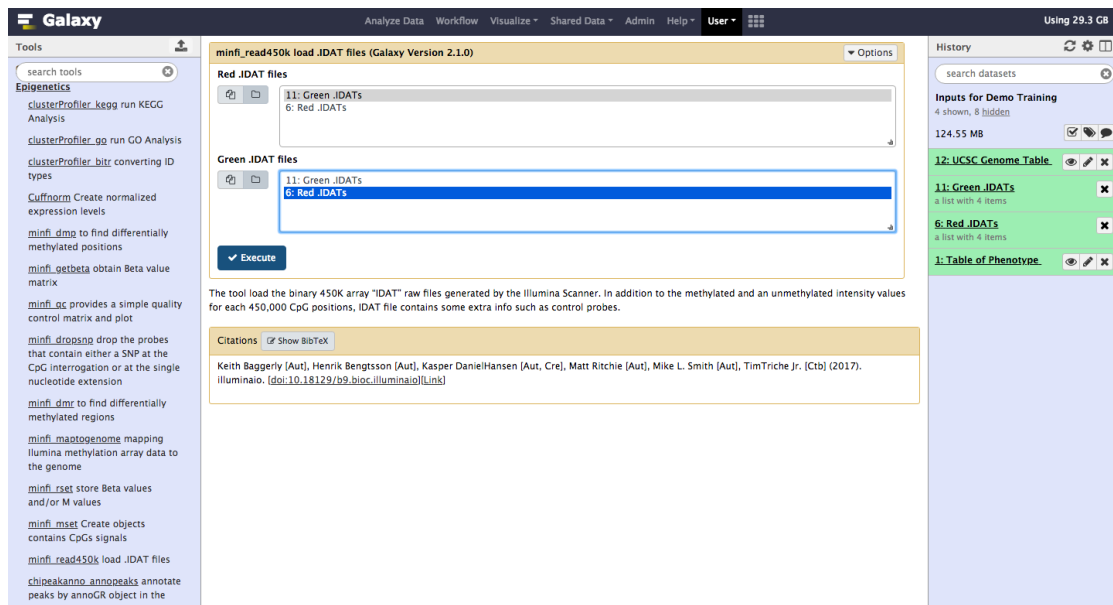


Figure A.1: Data Upload minfi_read450k Tool Interface

and uMed which are the medians of MethylSet signals (Meth and Unmeth). Comparing them against one another allows user to detect and remove low-quality samples A.2

1. Run minfi_mset to create MethylSet object
2. Run minfi_qc to estimate sample-specific quality control
3. Convert methylation data from the MethylSet , to ratios with minfi_rset
4. Then map ratio data to the genome using minfi_maptogenome tool

Removing probes affected by genetic variation

1. Run minfi_dropsnp to remove the probes that contain either a SNP at the metylated loci interrogation or at the single nucleotide extension, highly recommended by (Aryee et al., 2014)

DMPs and DMRs Identification

The main goal of the EWAS suite is to simplify the way differentially methylated loci sites are detected. The EWAS suite contains minfi_dmp tool detecting differentially methylated positions (DMPs) with respect to a phenotype covariate, and more complex minfi_dmr solution for finding differentially methylated regions (DMRs). Genomic regions that are differentially methylated between two conditions can be tracked using a bump hunting algorithm. The algorithm first implements a t-statistic at each methylated loci location,

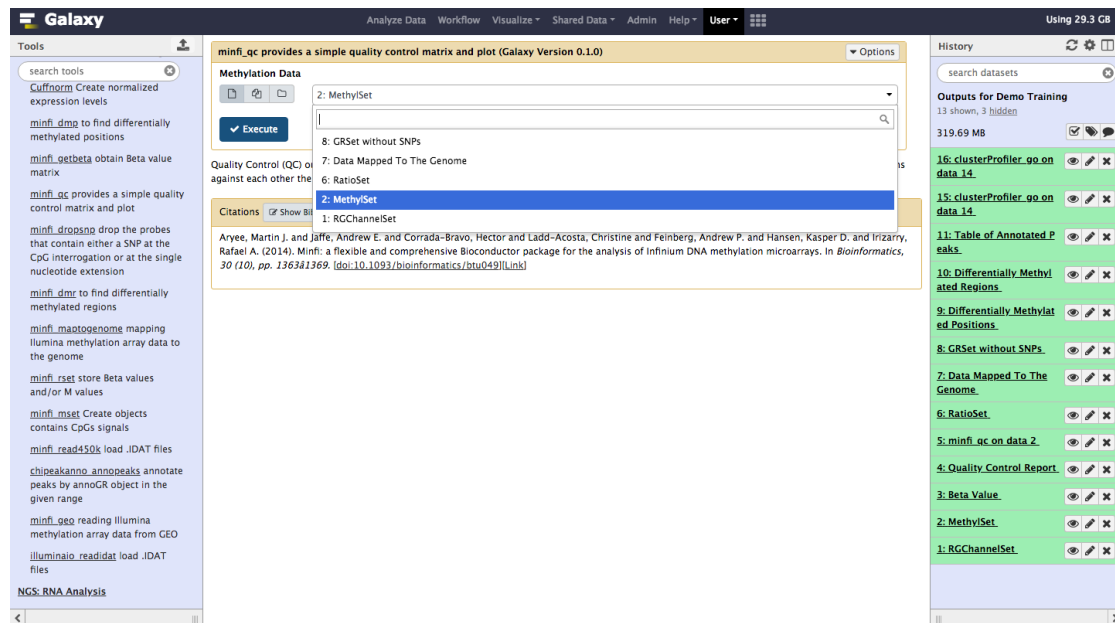


Figure A.2: Quality Assessment (minfi_qc) Tool Interface

with optional smoothing, then groups probes into clusters with a maximum location gap and a cutoff size.

1. Import phenotypeTable.txt from Zenodo ()
2. Run minfi dmp with the following parameters

Input set: GenomicRatioSet

Phenotype Table: phenotypeTable.txt

Phenotype Type: categorical

qCutoff Size: 0.5 (DMPs with an FDR q-value greater than this will not be returned)

Variance Shrinkage: TRUE (is recommended when sample sizes are small <10)

3. Run minfi_dmr A.3

Input set: GenomicRatioSet

Phenotype Table: phenotypeTable.txt

factor1: sensitive

factor2: resistant

maxGap Size: 250

coef Size: 2

Cutoff Size: 0.1

`nullMethod:permutation`

`verbose:TRUE`

4. Visualize Differentially Methylated Regions with UCSC

- Click on the minfi_dmr output in your history to expand it
- Click on the pencil button displayed in your dataset in the history set
- Database/Build Human Feb. 2009 (GRCh37/hg19) (hg19)
- Press Save
- Towards the bottom of the history item, find the line starting with display at UCSC
- This will launch UCSC Genome Browser (Kent et al., 2002) with your Custom Track A.4

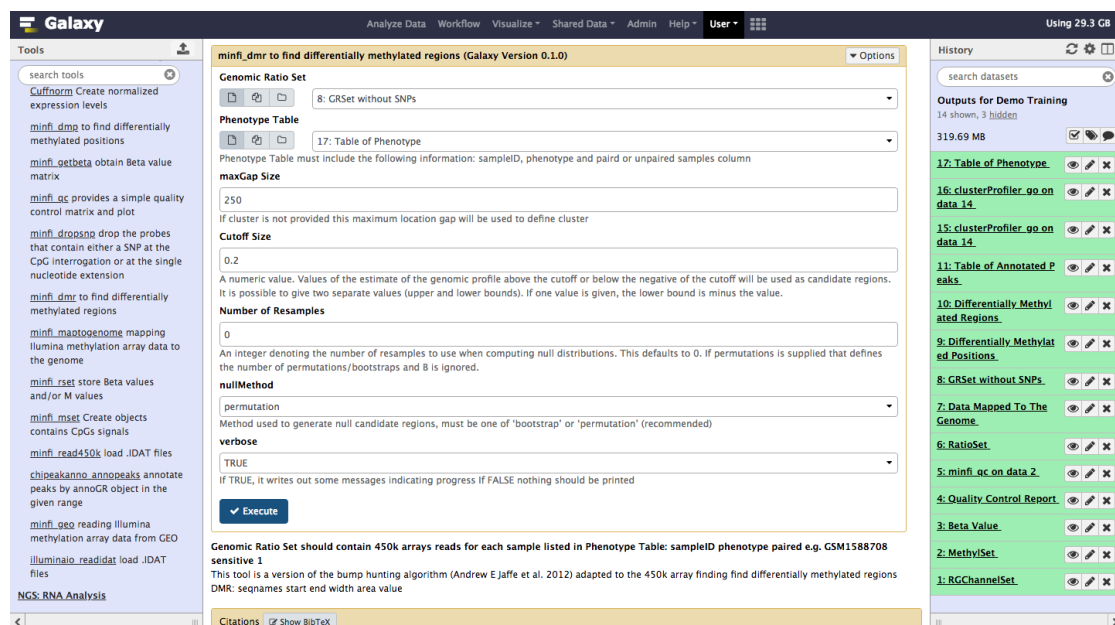


Figure A.3: minfi_dmr Tool Interface

Annotation and Visualization

In addition to downstream analysis users can annotate the differentially methylated loci to the promoter regions of genes with gene function description, and relationships between these concepts.

1. Run chipeakanno_anno peaks with the following parameters

Differentialy methylated data: use output of Differentially Methylated Positions analysis from Step 4

bindingType: StartSite

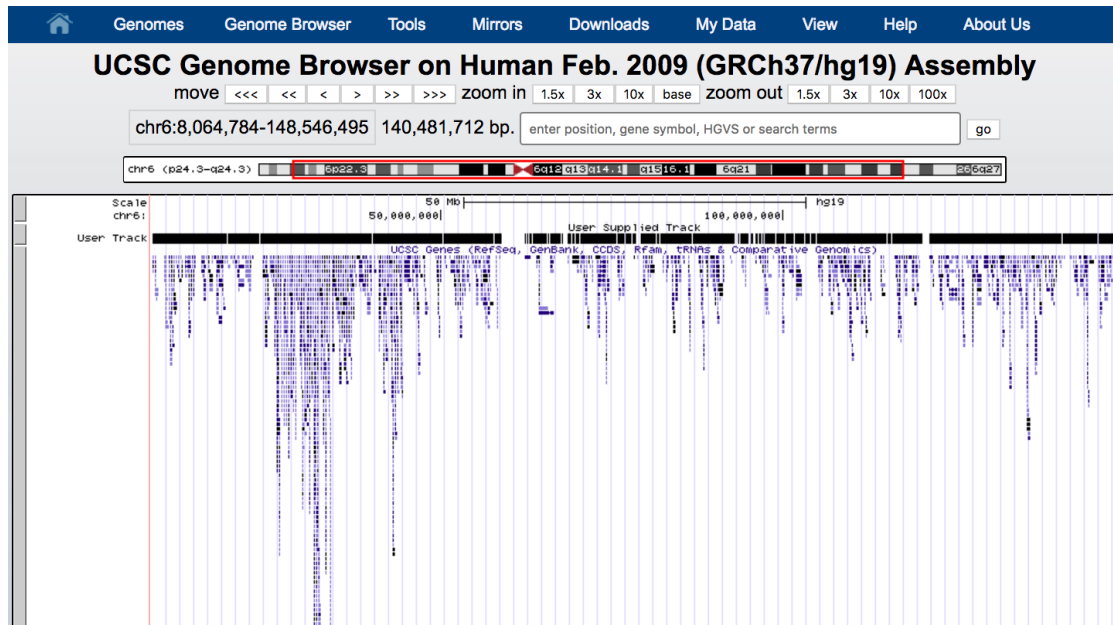


Figure A.4: DMR Training Track launch UCSC Genome Browser

bindingRegionStart: -5000

bindingRegionEnd: 3000

Additional Column of Score: 8 position of column of score optional value if it is required

2. Cut *gene_name* Column from Table of Annotated Peaks to get List of Genes with the following parameters

Cut columns: c16

Delimited by: Tab

3. Remove beginning of Gene List with the following parameter

Remove first: 1

4. Convert List of Genes to List of entrez ID using clusterProfiler bitr with the following parameters

Input Type Gene ID: SYMBOL

Output Type Gene ID: ENTREZID

5. Run GO Enrichment Analysis with clusterProfiler go using output of step 4

Appendix B

Table of Differentially Methylated Regions

Supplementary Data File

Description: The accompanying table shows the Differentially Methylated Regions founded in melanoma biopsies pre and post MAPKi treatment. The start and end columns indicate the limiting genomic locations of the DMR with width beetwen them. The value column indicates the average difference in methylation in the methylated region, and the area column indicates the area of this region with respect to the 0 line.

<i>seqnames</i>	<i>start</i>	<i>end</i>	<i>width</i>	<i>value</i>
<i>chr11</i>	637035	637175	141	−0.124012888048672
<i>chr6</i>	28584155	28584172	18	−0.10798311230643801
<i>chr13</i>	20710941	20711042	102	−0.103599195678859
<i>chr16</i>	14403004	14403022	19	−0.14040893898075399
<i>chr8</i>	11560486	11560510	25	−0.136436552550316
<i>chr2</i>	225307070	225307259	190	−0.13627480203925599
<i>chr3</i>	119421667	119421868	202	−0.129286989416137
<i>chr7</i>	2143886	2143942	57	0.12731922331307499
<i>chr14</i>	95235402	95235489	88	−0.12657979944939499
<i>chr5</i>	74907592	74907694	103	−0.12238116743333401
<i>chr8</i>	11539320	11539405	86	−0.118715555887448
<i>chr5</i>	180632948	180633063	116	0.117538987740608
<i>chr16</i>	1521617	1521656	40	0.116705694748048
<i>chr6</i>	33255241	33255400	160	0.115841009579616

<i>chr17</i>	77751069	77751089	21	0.11544281219399299
<i>chr3</i>	49027210	49027225	16	−0.11447089631264699
<i>chr6</i>	17282333	17282354	22	0.11134776487086499
<i>chr12</i>	14927345	14927351	7	−0.109710621025649
<i>chr4</i>	76555772	76555777	6	−0.10960929761368
<i>chr11</i>	63258744	63258779	36	−0.108360433208011
<i>chr22</i>	46688823	46688823	1	0.16053500897133899
<i>chr19</i>	55889216	55889216	1	0.15305514017642999
<i>chr11</i>	2292000	2292000	1	−0.14823121136736001
<i>chr1</i>	2987645	2987645	1	−0.14733788727401501
<i>chr12</i>	12008666	12008666	1	0.14329740027684401
<i>chr12</i>	66134770	66134770	1	0.13955551276179201
<i>chrX</i>	82763706	82763706	1	−0.13885729522767701
<i>chr6</i>	117584665	117584665	1	−0.13674957088845799
<i>chr12</i>	114845868	114845868	1	−0.134920614463754
<i>chr8</i>	141370229	141370229	1	0.134914365547156
<i>chr6</i>	170553845	170553845	1	0.13420248495128001
<i>chr2</i>	222435351	222435351	1	0.13203352052036099
<i>chr1</i>	230406371	230406371	1	0.13160625663473399
<i>chr4</i>	1623883	1623883	1	0.13095355759851199
<i>chr17</i>	37024625	37024625	1	0.130321842759
<i>chr14</i>	65007512	65007512	1	−0.130098860102633
<i>chrX</i>	132548278	132548278	1	0.12832582289404099
<i>chr6</i>	13326842	13326842	1	0.12831656980001699
<i>chr2</i>	242756362	242756362	1	0.12764825400728799
<i>chr14</i>	57274763	57274763	1	−0.12735053095825299

<i>chr10</i>	62761575	62761575	1	0.12726090174218599
<i>chrX</i>	48980610	48980610	1	-0.12688506472781799
<i>chr15</i>	59729553	59729553	1	0.12685394191065899
<i>chr17</i>	81014667	81014667	1	0.12622185598444399
<i>chr6</i>	28603779	28603779	1	-0.126196570119602
<i>chr10</i>	115999174	115999174	1	-0.12618908061598799
<i>chr19</i>	58220370	58220370	1	-0.12603239674417299
<i>chr2</i>	242756029	242756029	1	0.12537568512165301
<i>chr7</i>	35301188	35301188	1	-0.12514987300042599
<i>chr17</i>	26554610	26554610	1	-0.12463814819647499
<i>chr5</i>	134370282	134370282	1	0.124071381395868
<i>chr13</i>	111521981	111521981	1	0.123596167075517
<i>chr7</i>	27161749	27161749	1	-0.123318125974862
<i>chr3</i>	177315882	177315882	1	0.122654143071462
<i>chr6</i>	33255172	33255172	1	0.12215916290196301
<i>chr3</i>	137301727	137301727	1	0.1218114078845
<i>chr20</i>	30640256	30640256	1	-0.121679439121747
<i>chr21</i>	27945125	27945125	1	-0.121515821155445
<i>chr12</i>	4398508	4398508	1	0.12149381802154
<i>chr8</i>	142428240	142428240	1	-0.12104738724242201
<i>chr11</i>	68081226	68081226	1	0.12077674636437
<i>chr2</i>	242830161	242830161	1	0.120644928832901
<i>chr10</i>	88718317	88718317	1	0.12055366697212699

<i>chr8</i>	120868748	120868748	1	0.12002641277069299
<i>chr13</i>	25085301	25085301	1	-0.119932776811815
<i>chrX</i>	27999538	27999538	1	-0.119824688825988
<i>chr3</i>	157261021	157261021	1	-0.11974114896269
<i>chr7</i>	27224873	27224873	1	0.11970903424680999
<i>chr9</i>	117372723	117372723	1	0.119031262957996
<i>chr12</i>	124144432	124144432	1	-0.118915955262726
<i>chr1</i>	50884480	50884480	1	-0.11824224544209
<i>chr3</i>	49027243	49027243	1	-0.11756965654969299
<i>chr11</i>	72851913	72851913	1	0.11752549983777701
<i>chr10</i>	16562626	16562626	1	-0.11743482801303599
<i>chr2</i>	51259807	51259807	1	-0.117328845913815
<i>chr19</i>	15360953	15360953	1	-0.11666845196612199
<i>chr7</i>	6199980	6199980	1	0.116654583278869
<i>chr6</i>	33041218	33041218	1	-0.11631137363175
<i>chr17</i>	8926158	8926158	1	-0.11597704742722401
<i>chr4</i>	41883164	41883164	1	-0.115736384202522
<i>chr4</i>	154710750	154710750	1	-0.115709145966333
<i>chr21</i>	44851244	44851244	1	-0.115289672564198
<i>chr16</i>	85659946	85659946	1	0.11523346935845399
<i>chr3</i>	8671361	8671361	1	-0.115233017453939
<i>chr8</i>	65281496	65281496	1	-0.11476238720759301
<i>chrX</i>	68727150	68727150	1	-0.114706822290637
<i>chr8</i>	55380185	55380185	1	-0.114666685582534
<i>chr11</i>	6953558	6953558	1	-0.114601482040839
<i>chr16</i>	54967389	54967389	1	-0.11454073849728499
<i>chr6</i>	27235843	27235843	1	-0.114349696886029
<i>chr12</i>	1948648	1948648	1	-0.11432918842155999
<i>chr20</i>	30640022	30640022	1	-0.114211058386022
<i>chr5</i>	148810177	148810177	1	0.11387892996193701

<i>chr7</i>	47344608	47344608	1	0.113712054963099
<i>chr1</i>	22141170	22141170	1	-0.11351137352818
<i>chr3</i>	121741209	121741209	1	-0.113343572230443
<i>chr1</i>	2524890	2524890	1	0.113241054936184
<i>chr2</i>	242008100	242008100	1	0.113094021200255
<i>chr1</i>	3332188	3332188	1	0.113053786666189
<i>chr6</i>	170554795	170554795	1	0.112955245971986
<i>chr2</i>	220117771	220117771	1	0.112940541389401
<i>chr1</i>	13910667	13910667	1	-0.112356385525081
<i>chr11</i>	130271817	130271817	1	0.11223967583349601
<i>chr4</i>	164088478	164088478	1	0.112205903170614
<i>chr1</i>	874697	874697	1	0.112158136613177
<i>chrX</i>	30327778	30327778	1	0.11213589444989901
<i>chrX</i>	134305728	134305728	1	-0.111928225752649
<i>chr11</i>	60225240	60225240	1	-0.11165621383909299
<i>chr14</i>	97059005	97059005	1	-0.111652075228202
<i>chr7</i>	73803588	73803588	1	0.11148853924242901
<i>chr6</i>	28584103	28584103	1	-0.11147170176575701
<i>chr2</i>	160761085	160761085	1	-0.111359586185865
<i>chrX</i>	138276972	138276972	1	-0.111340640997727
<i>chr17</i>	17929033	17929033	1	-0.11023598405450299
<i>chr7</i>	15726466	15726466	1	-0.110191504951083
<i>chr7</i>	33080496	33080496	1	-0.110077665461544

<i>chr15</i>	33010399	33010399	1	−0.110006567837051
<i>chr6</i>	158834216	158834216	1	0.10999645741552
<i>chrX</i>	114957471	114957471	1	−0.10999201454742399
<i>chr6</i>	27513479	27513479	1	−0.109987017774902
<i>chr18</i>	44527026	44527026	1	−0.10998635677157299
<i>chr1</i>	1110107	1110107	1	0.109893101979087
<i>chr6</i>	33041343	33041343	1	−0.109868885925718
<i>chr10</i>	16562998	16562998	1	−0.10977522125084301
<i>chr3</i>	157261106	157261106	1	−0.10963935043367801
<i>chr6</i>	32920962	32920962	1	−0.10942442445153699
<i>chr12</i>	125034283	125034283	1	0.109238850116584
<i>chr5</i>	172671526	172671526	1	−0.109232336122665
<i>chr4</i>	24913973	24913973	1	−0.109168717974375
<i>chr20</i>	30640121	30640121	1	−0.109007215333146
<i>chr12</i>	103355958	103355958	1	−0.108909828259471
<i>chr17</i>	78955599	78955599	1	0.10888949659701599
<i>chr7</i>	1563708	1563708	1	0.10865079819256
<i>chr11</i>	20181725	20181725	1	−0.10859791759907
<i>chr4</i>	81189927	81189927	1	−0.108480853512708
<i>chr9</i>	100614879	100614879	1	−0.108460640192918
<i>chr10</i>	729226	729226	1	−0.108418232846721
<i>chr11</i>	32449821	32449821	1	−0.108412173538479
<i>chrX</i>	151620975	151620975	1	−0.108322288361311
<i>chr10</i>	127569905	127569905	1	0.108221794436341
<i>chr17</i>	6024017	6024017	1	−0.108161056408916
<i>chr15</i>	39871923	39871923	1	−0.108012224702111
<i>chr1</i>	16553549	16553549	1	−0.10800653773188899

<i>chr1</i>	1286917	1286917	1	0.107997667928319
<i>chr17</i>	63555244	63555244	1	-0.10793463833573499
<i>chr5</i>	758981	758981	1	-0.107887747828479
<i>chr3</i>	119832292	119832292	1	-0.107759788699141
<i>chr1</i>	119527638	119527638	1	-0.107720105666243
<i>chr7</i>	1883139	1883139	1	0.10760821356136401
<i>chr10</i>	43332443	43332443	1	-0.107599168427578
<i>chr7</i>	27232412	27232412	1	-0.10749113418684
<i>chr3</i>	73671377	73671377	1	-0.10747876722784699
<i>chr7</i>	96652481	96652481	1	0.107441203156391
<i>chr4</i>	55095209	55095209	1	0.107429344813642
<i>chr7</i>	104444687	104444687	1	0.107334363324657
<i>chr9</i>	115835616	115835616	1	-0.107318497484808
<i>chr7</i>	156736159	156736159	1	0.10730918740216799
<i>chr10</i>	73565625	73565625	1	0.107275081561613
<i>chr20</i>	9489749	9489749	1	-0.10726681239973899
<i>chr12</i>	133179338	133179338	1	0.10719873147812201
<i>chr19</i>	18497143	18497143	1	0.107180741689341
<i>chr22</i>	49067022	49067022	1	0.107160453820571
<i>chr7</i>	14025907	14025907	1	-0.107155044831899
<i>chr14</i>	101193038	101193038	1	-0.10705750690709399
<i>chr15</i>	75248086	75248086	1	0.10701489299339401
<i>chr16</i>	88499083	88499083	1	0.106969491433567
<i>chr17</i>	9080989	9080989	1	0.106864994393845
<i>chr18</i>	44526743	44526743	1	-0.106855303542586
<i>chr1</i>	2425888	2425888	1	0.106787124473451

<i>chr1</i>	200993200	200993200	1	0.10670070046810499
<i>chr21</i>	45811432	45811432	1	-0.10664739344672999
<i>chr8</i>	65291513	65291513	1	-0.106642175623474
<i>chr2</i>	19561482	19561482	1	0.106639719481052
<i>chr12</i>	124991139	124991139	1	0.10642104197168401
<i>chr7</i>	15726479	15726479	1	-0.106407197379582
<i>chr17</i>	57918500	57918500	1	-0.106337745838559
<i>chr7</i>	1234925	1234925	1	0.10630299054127
<i>chr1</i>	13910700	13910700	1	-0.106242503306801
<i>chr7</i>	2144579	2144579	1	0.10614186866653499
<i>chr16</i>	3088480	3088480	1	0.106020719334687
<i>chr6</i>	27526256	27526256	1	-0.105985650992152
<i>chr16</i>	4013337	4013337	1	0.105873126230512
<i>chr17</i>	48546193	48546193	1	-0.105780580885399
<i>chr8</i>	145013728	145013728	1	-0.10572300881592001
<i>chr10</i>	131641580	131641580	1	0.105659833057068
<i>chr10</i>	81946545	81946545	1	0.10545283207622599
<i>chr6</i>	11711971	11711971	1	-0.105382017269966
<i>chr17</i>	38984421	38984421	1	-0.10533591587878099
<i>chr14</i>	57265910	57265910	1	-0.105262348390998
<i>chr5</i>	140800929	140800929	1	-0.105229756006415
<i>chr18</i>	30352975	30352975	1	0.105227187967704
<i>chr12</i>	101603453	101603453	1	-0.105161186803192
<i>chr13</i>	102105440	102105440	1	0.105060953675732

<i>chr11</i>	32009163	32009163	1	−0.105029431478791
<i>chr6</i>	28477973	28477973	1	−0.10498215017005399
<i>chr19</i>	17610751	17610751	1	0.104948452961803
<i>chr1</i>	172886544	172886544	1	0.10491614248061799
<i>chr11</i>	68417376	68417376	1	−0.104905039443941
<i>chr3</i>	172166242	172166242	1	−0.104886809586504
<i>chr6</i>	170479840	170479840	1	0.10488226318586601
<i>chr7</i>	27184030	27184030	1	0.104859995048034
<i>chr8</i>	24772137	24772137	1	−0.104854886848308
<i>chr16</i>	46919112	46919112	1	0.10480590801688799
<i>chr14</i>	59038939	59038939	1	−0.104713619798806
<i>chr4</i>	154710523	154710523	1	−0.104518164900059
<i>chr3</i>	89164223	89164223	1	0.10451440965561699
<i>chr11</i>	111250093	111250093	1	−0.10451297615451501
<i>chr15</i>	81426610	81426610	1	−0.104499752568863
<i>chr14</i>	75153307	75153307	1	0.104490964925798
<i>chr2</i>	222333289	222333289	1	0.10435832104432299
<i>chr1</i>	241519652	241519652	1	−0.104279302631394
<i>chrX</i>	103268309	103268309	1	−0.104247652076186
<i>chr10</i>	28288192	28288192	1	−0.104246364439679
<i>chr4</i>	182862370	182862370	1	0.104243531040668
<i>chr11</i>	88069169	88069169	1	0.104212710647739
<i>chr8</i>	98290229	98290229	1	−0.10418878749230701
<i>chr14</i>	62068941	62068941	1	0.10416661372495301

<i>chr6</i>	33255541	33255541	1	0.10403223430929
<i>chr19</i>	595705	595705	1	0.10400729134155599
<i>chr12</i>	13043728	13043728	1	0.103865906867918
<i>chr17</i>	74260692	74260692	1	−0.103836859804503
<i>chr17</i>	75276428	75276428	1	0.10382569025496
<i>chr12</i>	66123127	66123127	1	−0.10382279315894601
<i>chr15</i>	25414716	25414716	1	−0.10370450501645
<i>chr7</i>	1095005	1095005	1	0.103665737875729
<i>chr14</i>	91752093	91752093	1	−0.103659140301418
<i>chr3</i>	172166517	172166517	1	−0.10361076137165701
<i>chr16</i>	1076283	1076283	1	0.1034982175643
<i>chr8</i>	37557348	37557348	1	0.103494253879458
<i>chr7</i>	51544475	51544475	1	−0.103484496718387
<i>chr6</i>	159084599	159084599	1	0.10346642150404201
<i>chr2</i>	11672761	11672761	1	−0.10301523610233899
<i>chr2</i>	172958324	172958324	1	0.102865114465247
<i>chr19</i>	18540330	18540330	1	0.102851604206926
<i>chr2</i>	222064243	222064243	1	0.10282718638032901
<i>chr6</i>	3054085	3054085	1	−0.102745208692728
<i>chr7</i>	92645767	92645767	1	0.10272223311284299
<i>chr8</i>	58056113	58056113	1	−0.102622883478428
<i>chr8</i>	17353980	17353980	1	0.10257945756098601
<i>chr20</i>	3229402	3229402	1	−0.102376606295041

<i>chr2</i>	242606122	242606122	1	0.102372151482944
<i>chr21</i>	40145361	40145361	1	−0.102349240990334
<i>chr10</i>	29186681	29186681	1	0.102347581984831
<i>chr1</i>	156612436	156612436	1	0.102330099226989
<i>chr10</i>	76727919	76727919	1	0.10232007528624699
<i>chr15</i>	96868971	96868971	1	0.10229919224851899
<i>chr19</i>	12306198	12306198	1	−0.102230147125593
<i>chr18</i>	67959378	67959378	1	0.102164030230496
<i>chr2</i>	182543233	182543233	1	−0.102125446159189
<i>chr11</i>	122612858	122612858	1	0.102097752801121
<i>chr3</i>	49027156	49027156	1	−0.102058470880869
<i>chr12</i>	41087520	41087520	1	0.101988875236978
<i>chr1</i>	92942075	92942075	1	0.101985251286454
<i>chr8</i>	65281397	65281397	1	−0.10197902083079299
<i>chr8</i>	110988071	110988071	1	−0.10189553138667699
<i>chr18</i>	59483441	59483441	1	−0.10189418183363901
<i>chr2</i>	219857231	219857231	1	0.101866327505896
<i>chr7</i>	48129904	48129904	1	−0.10179139338468
<i>chr10</i>	70847350	70847350	1	−0.101470034442445
<i>chr1</i>	47882322	47882322	1	−0.101353260878196
<i>chr10</i>	504497	504497	1	0.101307091968666
<i>chr7</i>	1233469	1233469	1	0.101251969857042

<i>chr3</i>	26663593	26663593	1	−0.101207719948323
<i>chr12</i>	124990897	124990897	1	0.101196428120161
<i>chrX</i>	153770983	153770983	1	−0.101179944369073
<i>chrY</i>	9385586	9385586	1	−0.101155812375696
<i>chr10</i>	129861770	129861770	1	0.10110526762918701
<i>chr10</i>	80523167	80523167	1	−0.10109858573823299
<i>chr22</i>	49137906	49137906	1	0.101094556847243
<i>chr15</i>	93633408	93633408	1	0.10108409025040201
<i>chr2</i>	163695776	163695776	1	−0.100983554882114
<i>chr14</i>	57275967	57275967	1	−0.100920098537103
<i>chr6</i>	42738967	42738967	1	−0.10087877674661901
<i>chr6</i>	27258466	27258466	1	−0.100870954020625
<i>chr13</i>	114086994	114086994	1	0.100799446744015
<i>chr10</i>	85997177	85997177	1	−0.100700692309193
<i>chr4</i>	122686038	122686038	1	−0.100674866753929
<i>chr16</i>	10449523	10449523	1	−0.100622143068377
<i>chr16</i>	82044957	82044957	1	−0.10056489087438
<i>chr11</i>	6913644	6913644	1	−0.100467407127736
<i>chr6</i>	27342632	27342632	1	−0.100280738631745
<i>chr6</i>	111888446	111888446	1	0.100264975037983
<i>chr14</i>	59742285	59742285	1	−0.1002495184215
<i>chr12</i>	53227816	53227816	1	−0.10022075519222599
<i>chr6</i>	10398601	10398601	1	−0.100188920898861
<i>chr11</i>	32458714	32458714	1	−0.100119136435286
<i>chr17</i>	37894397	37894397	1	0.10004974772408801
<i>chr4</i>	4857506	4857506	1	0.10003454190038

Appendix C

Table of DMR's Annotation

Supplementary Data File

Description: The accompanying tables shows the annotation of DMR's founded in melanoma biopsies pre and post MAPKi treatment.

<i>seqnames</i>	<i>start</i>	<i>end</i>	<i>width</i>	<i>gene_name</i>
<i>chr11</i>	637035	637175	141	<i>DRD4</i>
<i>chr6</i>	28584155	28584172	18	<i>SCAND3</i>
<i>chr16</i>	14403004	14403022	19	<i>MIR365A</i>
<i>chr3</i>	119421667	119421868	202	<i>MAATS1</i>
<i>chr14</i>	95235402	95235489	88	<i>GSC</i>
<i>chr5</i>	74907592	74907694	103	<i>ANKDD1B</i>
<i>chr5</i>	180632948	180633063	116	<i>TRIM7</i>
<i>chr5</i>	180632948	180633063	116	<i>CTC – 338M12.1</i>
<i>chr16</i>	1521617	1521656	40	<i>LA16c – 390E6.3</i>
<i>chr6</i>	33255241	33255400	160	<i>WDR46</i>
<i>chr6</i>	33255241	33255400	160	<i>PFDN6</i>
<i>chr17</i>	77751069	77751089	21	<i>CBX2</i>
<i>chr3</i>	49027210	49027225	16	<i>RP13 – 131K19.2</i>
<i>chr3</i>	49027210	49027225	16	<i>RP13 – 131K19.7</i>

<i>chr3</i>	49027210	49027225	16	<i>P4HTM</i>
<i>chr6</i>	17282333	17282354	22	<i>RBM24</i>
<i>chr12</i>	14927345	14927351	7	<i>HIST4H4</i>
<i>chr12</i>	14927345	14927351	7	<i>H2AFJ</i>
<i>chr4</i>	76555772	76555777	6	<i>CDKL2</i>
<i>chr11</i>	63258744	63258779	36	<i>HRASLS5</i>
<i>chr22</i>	46688823	46688823	1	<i>GTSE1</i>
<i>chr19</i>	55889216	55889216	1	<i>TMEM190</i>
<i>chr19</i>	55889216	55889216	1	<i>CTD – 2105E13.15</i>
<i>chr11</i>	2292000	2292000	1	<i>ASCL2</i>
<i>chr1</i>	2987645	2987645	1	<i>LINC00982</i>
<i>chr1</i>	2987645	2987645	1	<i>PRDM16</i>
<i>chrX</i>	82763706	82763706	1	<i>RP3 – 326L13.2</i>
<i>chrX</i>	82763706	82763706	1	<i>POU3F4</i>
<i>chrX</i>	82763706	82763706	1	<i>RP3 – 326L13.3</i>
<i>chr6</i>	117584665	117584665	1	<i>VGLL2</i>
<i>chr12</i>	114845868	114845868	1	<i>TBX5</i>
<i>chr12</i>	114845868	114845868	1	<i>TBX5 – AS1</i>
<i>chr2</i>	222435351	222435351	1	<i>CTD – 2308L22.1</i>
<i>chr1</i>	230406371	230406371	1	<i>RP5 – 956O18.2</i>
<i>chr17</i>	37024625	37024625	1	<i>LASP1</i>
<i>chr14</i>	65007512	65007512	1	<i>RP11 – 973N13.4</i>
<i>chr14</i>	65007512	65007512	1	<i>RP11 – 973N13.5</i>
<i>chrX</i>	132548278	132548278	1	<i>GPC4</i>
<i>chr6</i>	13326842	13326842	1	<i>TBC1D7</i>
<i>chr2</i>	242756362	242756362	1	<i>AC114730.3</i>
<i>chr14</i>	57274763	57274763	1	<i>OTX2</i>

<i>chr10</i>	62761575	62761575	1	<i>RHOBTB1</i>
<i>chrX</i>	48980610	48980610	1	<i>GPKOW</i>
<i>chr17</i>	81014667	81014667	1	<i>B3GNTL1</i>
<i>chr6</i>	28603779	28603779	1	<i>RP11 – 373N24.2</i>
<i>chr10</i>	115999174	115999174	1	<i>VWA2</i>
<i>chr19</i>	58220370	58220370	1	<i>ZNF154</i>
<i>chr2</i>	242756029	242756029	1	<i>AC114730.3</i>
<i>chr7</i>	35301188	35301188	1	<i>AC009531.2</i>
<i>chr17</i>	26554610	26554610	1	<i>PYY2</i>
<i>chr5</i>	134370282	134370282	1	<i>PITX1</i>
<i>chr5</i>	134370282	134370282	1	<i>C5orf66</i>
<i>chr13</i>	111521981	111521981	1	<i>LINC00346</i>
<i>chr6</i>	33255172	33255172	1	<i>WDR46</i>
<i>chr6</i>	33255172	33255172	1	<i>PFDN6</i>
<i>chr20</i>	30640256	30640256	1	<i>HCK</i>
<i>chr21</i>	27945125	27945125	1	<i>CYYR1</i>
<i>chr11</i>	68081226	68081226	1	<i>LRP5</i>
<i>chr10</i>	88718317	88718317	1	<i>SNCG</i>
<i>chr8</i>	120868748	120868748	1	<i>DSCC1</i>
<i>chr13</i>	25085301	25085301	1	<i>PARP4</i>
<i>chrX</i>	27999538	27999538	1	<i>DCAF8L1</i>
<i>chr3</i>	157261021	157261021	1	<i>C3orf55</i>
<i>chr7</i>	27224873	27224873	1	<i>HOXA10</i>
<i>chr7</i>	27224873	27224873	1	<i>HOXA11</i>
<i>chr7</i>	27224873	27224873	1	<i>HOXA11 – AS</i>
<i>chr9</i>	117372723	117372723	1	<i>C9orf91</i>

<i>chr3</i>	49027243	49027243	1	<i>RP13 – 131K19.2</i>
<i>chr3</i>	49027243	49027243	1	<i>RP13 – 131K19.7</i>
<i>chr3</i>	49027243	49027243	1	<i>P4HTM</i>
<i>chr11</i>	72851913	72851913	1	<i>FCHSD2</i>
<i>chr10</i>	16562626	16562626	1	<i>C1QL3</i>
<i>chr2</i>	51259807	51259807	1	<i>NRXN1</i>
<i>chr2</i>	51259807	51259807	1	<i>AC007682.1</i>
<i>chr6</i>	33041218	33041218	1	<i>HLA – DPB1</i>
<i>chr17</i>	8926158	8926158	1	<i>NTN1</i>
<i>chr4</i>	41883164	41883164	1	<i>LINC00682</i>
<i>chr4</i>	154710750	154710750	1	<i>SFRP2</i>
<i>chr21</i>	44851244	44851244	1	<i>SIK1</i>
<i>chr8</i>	65281496	65281496	1	<i>RP11 – 32K4.1</i>
<i>chr8</i>	65281496	65281496	1	<i>LINC00966</i>
<i>chrX</i>	68727150	68727150	1	<i>FAM155B</i>
<i>chr16</i>	54967389	54967389	1	<i>CRNDE</i>
<i>chr16</i>	54967389	54967389	1	<i>IRX5</i>
<i>chr16</i>	54967389	54967389	1	<i>CTD – 3032H12.2</i>
<i>chr6</i>	27235843	27235843	1	<i>XXbac – BPGBPG24O18.1</i>
<i>chr20</i>	30640022	30640022	1	<i>HCK</i>
<i>chr5</i>	148810177	148810177	1	<i>MIR145</i>
<i>chr1</i>	22141170	22141170	1	<i>LDLRAD2</i>
<i>chr3</i>	121741209	121741209	1	<i>ILDR1</i>
<i>chr2</i>	242008100	242008100	1	<i>AC005237.4</i>

<i>chr2</i>	220117771	220117771	1	<i>TUBA4B</i>
<i>chr1</i>	13910667	13910667	1	<i>PDPN</i>
<i>chr4</i>	164088478	164088478	1	<i>NAF1</i>
<i>chrX</i>	30327778	30327778	1	<i>NR0B1</i>
<i>chrX</i>	134305728	134305728	1	<i>CT55</i>
<i>chr11</i>	60225240	60225240	1	<i>MS4A1</i>
<i>chr11</i>	60225240	60225240	1	<i>LRG_140</i>
<i>chr14</i>	97059005	97059005	1	<i>RP11 – 433J8.1</i>
<i>chr6</i>	28584103	28584103	1	<i>SCAND3</i>
<i>chr2</i>	160761085	160761085	1	<i>LY75</i>
<i>chr2</i>	160761085	160761085	1	<i>LY75 – CD302</i>
<i>chr7</i>	15726466	15726466	1	<i>MEOX2</i>
<i>chr7</i>	15726466	15726466	1	<i>AC005550.4</i>
<i>chr7</i>	33080496	33080496	1	<i>AC074338.4</i>
<i>chr15</i>	33010399	33010399	1	<i>RP11 – 758N13.1</i>
<i>chr15</i>	33010399	33010399	1	<i>GREM1</i>
<i>chrX</i>	114957471	114957471	1	<i>RP1 – 241P17.4</i>
<i>chrX</i>	114957471	114957471	1	<i>AC005000.1</i>
<i>chrX</i>	114957471	114957471	1	<i>RP1 – 241P17.1</i>
<i>chr1</i>	1110107	1110107	1	<i>TTLL10</i>
<i>chr6</i>	33041343	33041343	1	<i>HLA – DPB1</i>
<i>chr10</i>	16562998	16562998	1	<i>C1QL3</i>
<i>chr3</i>	157261106	157261106	1	<i>C3orf55</i>

<i>chr6</i>	32920962	32920962	1	<i>XXbac – BPG181M17.5</i>
<i>chr20</i>	30640121	30640121	1	<i>HCK</i>
<i>chr12</i>	103355958	103355958	1	<i>PAH</i>
<i>chr11</i>	20181725	20181725	1	<i>DBX1</i>
<i>chr4</i>	81189927	81189927	1	<i>FGF5</i>
<i>chr9</i>	100614879	100614879	1	<i>FOX E1</i>
<i>chrX</i>	151620975	151620975	1	<i>GABRA3</i>
<i>chr15</i>	39871923	39871923	1	<i>THBS1</i>
<i>chr1</i>	16553549	16553549	1	<i>ANO7P1</i>
<i>chr1</i>	1286917	1286917	1	<i>DVL1</i>
<i>chr17</i>	63555244	63555244	1	<i>AXIN2</i>
<i>chr17</i>	63555244	63555244	1	<i>LRG_296</i>
<i>chr3</i>	119832292	119832292	1	<i>RN7SL762P</i>
<i>chr10</i>	43332443	43332443	1	<i>RNU6 – 885P</i>
<i>chr7</i>	27232412	27232412	1	<i>RP1 – 170O19.14</i>
<i>chr3</i>	73671377	73671377	1	<i>PDZRN3</i>
<i>chr3</i>	73671377	73671377	1	<i>PDZRN3 – AS1</i>
<i>chr7</i>	96652481	96652481	1	<i>DLX5</i>
<i>chr4</i>	55095209	55095209	1	<i>PDGFRA</i>
<i>chr7</i>	104444687	104444687	1	<i>LHFPL3 – AS1</i>
<i>chr7</i>	156736159	156736159	1	<i>RP5 – 1121A15.3</i>

<i>chr19</i>	18497143	18497143	1	<i>MIR3189</i>
<i>chr14</i>	101193038	101193038	1	<i>DLK1</i>
<i>chr15</i>	75248086	75248086	1	<i>RPP25</i>
<i>chr15</i>	75248086	75248086	1	<i>SCAMP5</i>
<i>chr17</i>	9080989	9080989	1	<i>RP11 – 85B7.2</i>
<i>chr1</i>	2425888	2425888	1	<i>RP3 – 395M20.2</i>
<i>chr1</i>	2425888	2425888	1	<i>RP3 – 395M20.3</i>
<i>chr1</i>	200993200	200993200	1	<i>KIF21B</i>
<i>chr1</i>	200993200	200993200	1	<i>RP11 – 168O16.1</i>
<i>chr8</i>	65291513	65291513	1	<i>MIR124 – 2</i>
<i>chr2</i>	19561482	19561482	1	<i>OSR1</i>
<i>chr7</i>	15726479	15726479	1	<i>MEOX2</i>
<i>chr7</i>	15726479	15726479	1	<i>AC005550.4</i>
<i>chr17</i>	57918500	57918500	1	<i>MIR21</i>
<i>chr1</i>	13910700	13910700	1	<i>PDPN</i>
<i>chr16</i>	3088480	3088480	1	<i>CCDC64B</i>
<i>chr17</i>	48546193	48546193	1	<i>CHAD</i>
<i>chr10</i>	131641580	131641580	1	<i>MIR4297</i>
<i>chr5</i>	140800929	140800929	1	<i>PCDHGA11</i>
<i>chr5</i>	140800929	140800929	1	<i>PCDHGB8P</i>
<i>chr18</i>	30352975	30352975	1	<i>KLHL14</i>
<i>chr12</i>	101603453	101603453	1	<i>SLC5A8</i>

<i>chr13</i>	102105440	102105440	1	<i>ITGBL1</i>
<i>chr3</i>	172166242	172166242	1	<i>GHSR</i>
<i>chr7</i>	27184030	27184030	1	<i>HOXA5</i>
<i>chr8</i>	24772137	24772137	1	<i>RP11 – 624C23.1</i>
<i>chr8</i>	24772137	24772137	1	<i>GS1 – 72M22.1</i>
<i>chr8</i>	24772137	24772137	1	<i>NEFM</i>
<i>chr16</i>	46919112	46919112	1	<i>GPT2</i>
<i>chr4</i>	154710523	154710523	1	<i>SFRP2</i>
<i>chr11</i>	111250093	111250093	1	<i>RNU2 – 60P</i>
<i>chr1</i>	241519652	241519652	1	<i>RGS7</i>
<i>chrX</i>	103268309	103268309	1	<i>H2BFWT</i>
<i>chrX</i>	103268309	103268309	1	<i>H2BFM</i>
<i>chr10</i>	28288192	28288192	1	<i>ARMC4</i>
<i>chr10</i>	28288192	28288192	1	<i>RP11 – 218D6.4</i>
<i>chr11</i>	88069169	88069169	1	<i>CTSC</i>
<i>chr11</i>	88069169	88069169	1	<i>LRG_50</i>
<i>chr8</i>	98290229	98290229	1	<i>TSPYL5</i>
<i>chr6</i>	33255541	33255541	1	<i>WDR46</i>
<i>chr6</i>	33255541	33255541	1	<i>PFDN6</i>
<i>chr19</i>	595705	595705	1	<i>AC005559.3</i>
<i>chr12</i>	13043728	13043728	1	<i>GPRC5A</i>
<i>chr17</i>	74260692	74260692	1	<i>UBALD2</i>

<i>chr17</i>	75276428	75276428	1	<i>RP11 – 285E9.6</i>
<i>chr17</i>	75276428	75276428	1	40057
<i>chr15</i>	25414716	25414716	1	<i>TMEM261P1</i>
<i>chr15</i>	25414716	25414716	1	<i>SNORD115 – 1</i>
<i>chr15</i>	25414716	25414716	1	<i>SNORD115 – 2</i>
<i>chr3</i>	172166517	172166517	1	<i>GHSR</i>
<i>chr2</i>	11672761	11672761	1	<i>GREB1</i>
<i>chr6</i>	3054085	3054085	1	<i>RP1 – 40E16.2</i>
<i>chr8</i>	58056113	58056113	1	<i>RP11 – 513O17.2</i>
<i>chr8</i>	17353980	17353980	1	<i>SLC7A2</i>
<i>chr1</i>	156612436	156612436	1	<i>RP11 – 284F21.9</i>
<i>chr1</i>	156612436	156612436	1	<i>BCAN</i>
<i>chr1</i>	156612436	156612436	1	<i>RP11 – 284F21.10</i>
<i>chr15</i>	96868971	96868971	1	<i>NR2F2 – AS1</i>
<i>chr15</i>	96868971	96868971	1	<i>NR2F2</i>
<i>chr19</i>	12306198	12306198	1	<i>CTD – 2666L21.1</i>
<i>chr2</i>	182543233	182543233	1	<i>CERKL</i>
<i>chr2</i>	182543233	182543233	1	<i>NEUROD1</i>
<i>chr2</i>	182543233	182543233	1	<i>AC013733.3</i>
<i>chr3</i>	49027156	49027156	1	<i>RP13 – 131K19.2</i>
<i>chr3</i>	49027156	49027156	1	<i>RP13 – 131K19.7</i>
<i>chr3</i>	49027156	49027156	1	<i>P4HTM</i>
<i>chr12</i>	41087520	41087520	1	<i>CNTN1</i>
<i>chr8</i>	65281397	65281397	1	<i>RP11 – 32K4.1</i>
<i>chr8</i>	65281397	65281397	1	<i>LINC00966</i>
<i>chr8</i>	110988071	110988071	1	<i>KCNV1</i>
<i>chr2</i>	219857231	219857231	1	<i>CRYBA2</i>
<i>chr7</i>	48129904	48129904	1	<i>UPP1</i>

<i>chr10</i>	70847350	70847350	1	<i>SRGN</i>
<i>chr1</i>	47882322	47882322	1	<i>FOX E3</i>
<i>chr3</i>	26663593	26663593	1	<i>AC099754.1</i>
<i>chr3</i>	26663593	26663593	1	<i>LRRC3B</i>
<i>chrX</i>	153770983	153770983	1	<i>IKBKG</i>
<i>chrY</i>	9385586	9385586	1	<i>FAM197Y1</i>
<i>chrY</i>	9385586	9385586	1	<i>TSPY15P</i>
<i>chr15</i>	93633408	93633408	1	<i>RGMA</i>
<i>chr2</i>	163695776	163695776	1	<i>KCNH7</i>
<i>chr14</i>	57275967	57275967	1	<i>OTX2</i>
<i>chr14</i>	57275967	57275967	1	<i>OTX2 – AS1</i>
<i>chr4</i>	122686038	122686038	1	<i>TMEM155</i>
<i>chr4</i>	122686038	122686038	1	<i>AC079341.1</i>
<i>chr16</i>	10449523	10449523	1	<i>RP11 – 609N14.1</i>
<i>chr16</i>	82044957	82044957	1	<i>SDR42E1</i>
<i>chr11</i>	6913644	6913644	1	<i>OR2D2</i>
<i>chr6</i>	27342632	27342632	1	<i>ZNF204P</i>
<i>chr6</i>	27342632	27342632	1	<i>ZNF391</i>
<i>chr12</i>	53227816	53227816	1	<i>KRT79</i>
<i>chr12</i>	53227816	53227816	1	<i>RP11 – 153F5.3</i>
<i>chr11</i>	32458714	32458714	1	<i>WT1</i>
<i>chr11</i>	32458714	32458714	1	<i>WT1 – AS</i>
<i>chr17</i>	37894397	37894397	1	<i>GRB7</i>
<i>chr4</i>	4857506	4857506	1	<i>MSX1</i>

<i>gene_name</i>	<i>score</i>	<i>distanceToSite</i>
<i>DRD4</i>	-0.124012888048672	117
<i>SCAND3</i>	-0.10798311230643801	165
<i>MIR365A</i>	-0.14040893898075399	119
<i>MAATS1</i>	-0.129286989416137	0
<i>GSC</i>	-0.12657979944939499	1072
<i>ANKDD1B</i>	-0.12238116743333401	307
<i>TRIM7</i>	0.117538987740608	654
<i>CTC – 338M12.1</i>	0.117538987740608	56
<i>LA16c – 390E6.3</i>	0.116705694748048	1028
<i>WDR46</i>	0.115841009579616	1903
<i>PFDN6</i>	0.115841009579616	1678
<i>CBX2</i>	0.11544281219399299	841
<i>RP13 – 131K19.2</i>	-0.11447089631264699	195
<i>RP13 – 131K19.7</i>	-0.11447089631264699	3813
<i>P4HTM</i>	-0.11447089631264699	93
<i>RBM24</i>	0.11134776487086499	755
<i>HIST4H4</i>	-0.109710621025649	3279
<i>H2AFJ</i>	-0.109710621025649	74
<i>CDKL2</i>	-0.10960929761368	122

<i>HRASLS5</i>	−0.108360433208011	77
<i>GTSE1</i>	0.16053500897133899	3814
<i>TMEM190</i>	0.15305514017642999	1011
<i>CTD – 2105E13.15</i>	0.15305514017642999	276
<i>ASCL2</i>	−0.14823121136736001	181
<i>LINC00982</i>	−0.14733788727401501	2643
<i>PRDM16</i>	−0.14733788727401501	1912
<i>RP3 – 326L13.2</i>	−0.13885729522767701	515
<i>POU3F4</i>	−0.13885729522767701	436
<i>RP3 – 326L13.3</i>	−0.13885729522767701	2597
<i>VGLL2</i>	−0.13674957088845799	2055
<i>TBX5</i>	−0.134920614463754	378
<i>TBX5 – AS1</i>	−0.134920614463754	127
<i>CTD – 2308L22.1</i>	0.13203352052036099	1874
<i>RP5 – 956O18.2</i>	0.13160625663473399	2141
<i>LASP1</i>	0.130321842759	1486
<i>RP11 – 973N13.4</i>	−0.130098860102633	425
<i>RP11 – 973N13.5</i>	−0.130098860102633	4914
<i>GPC4</i>	0.12832582289404099	1239
<i>TBC1D7</i>	0.12831656980001699	1972
<i>AC114730.3</i>	0.12764825400728799	3250
<i>OTX2</i>	−0.12735053095825299	2433
<i>RHOBTB1</i>	0.12726090174218599	376

<i>GPLOW</i>	−0.12688506472781799	458
<i>B3GNTL1</i>	0.12622185598444399	4980
<i>RP11 – 373N24.2</i>	−0.126196570119602	2407
<i>VWA2</i>	−0.12618908061598799	155
<i>ZNF154</i>	−0.12603239674417299	208
<i>AC114730.3</i>	0.12537568512165301	2917
<i>AC009531.2</i>	−0.12514987300042599	1847
<i>PYY2</i>	−0.12463814819647499	1020
<i>PITX1</i>	0.124071381395868	220
<i>C5orf66</i>	0.124071381395868	1311
<i>LINC00346</i>	0.123596167075517	180
<i>WDR46</i>	0.12215916290196301	2131
<i>PFDN6</i>	0.12215916290196301	1906
<i>HCK</i>	−0.121679439121747	264
<i>CYYR1</i>	−0.121515821155445	477
<i>LRP5</i>	0.12077674636437	1148
<i>SNCG</i>	0.12055366697212699	57
<i>DSCC1</i>	0.12002641277069299	497
<i>PARP4</i>	−0.119932776811815	1646
<i>DCAF8L1</i>	−0.119824688825988	27
<i>C3orf55</i>	−0.11974114896269	13
<i>HOXA10</i>	0.11970903424680999	4992
<i>HOXA11</i>	0.11970903424680999	30
<i>HOXA11 – AS</i>	0.11970903424680999	735

<i>C9orf91</i>	0.119031262957996	762
<i>RP13 – 131K19.2</i>	−0.11756965654969299	177
<i>RP13 – 131K19.7</i>	−0.11756965654969299	3846
<i>P4HTM</i>	−0.11756965654969299	75
<i>FCHSD2</i>	0.11752549983777701	1392
<i>C1QL3</i>	−0.11743482801303599	1377
<i>NRXN1</i>	−0.117328845913815	132
<i>AC007682.1</i>	−0.117328845913815	67
<i>HLA – DPB1</i>	−0.11631137363175	2484
<i>NTN1</i>	−0.11597704742722401	1298
<i>LINC00682</i>	−0.115736384202522	1463
<i>SFRP2</i>	−0.115709145966333	477
<i>SIK1</i>	−0.115289672564198	4235
<i>RP11 – 32K4.1</i>	−0.11476238720759301	380
<i>LINC00966</i>	−0.11476238720759301	4388
<i>FAM155B</i>	−0.114706822290637	2065
<i>CRNDE</i>	−0.11454073849728499	4287
<i>IRX5</i>	−0.11454073849728499	2614
<i>CTD – 3032H12.2</i>	−0.11454073849728499	1435
<i>XXbac – BPGBPG24O18.1</i>	−0.114349696886029	48
<i>HCK</i>	−0.114211058386022	30
<i>MIR145</i>	0.11387892996193701	327
<i>LDLRAD2</i>	−0.11351137352818	2411
<i>ILDR1</i>	−0.113343572230443	157

<i>AC005237.4</i>	0.113094021200255	4566
<i>TUBA4B</i>	0.112940541389401	193
<i>PDPN</i>	−0.112356385525081	706
<i>NAF1</i>	0.112205903170614	404
<i>NR0B1</i>	0.11213589444989901	62
<i>CT55</i>	−0.111928225752649	405
<i>MS4A1</i>	−0.11165621383909299	2014
<i>LRG_140</i>	−0.11165621383909299	1957
<i>RP11 – 433J8.1</i>	−0.111652075228202	64
<i>SCAND3</i>	−0.11147170176575701	113
<i>LY75</i>	−0.111359586185865	174
<i>LY75 – CD302</i>	−0.111359586185865	135
<i>MEOX2</i>	−0.110191504951083	28
<i>AC005550.4</i>	−0.110191504951083	1536
<i>AC074338.4</i>	−0.110077665461544	4501
<i>RP11 – 758N13.1</i>	−0.110006567837051	808
<i>GREM1</i>	−0.110006567837051	223
<i>RP1 – 241P17.4</i>	−0.10999201454742399	3801
<i>AC005000.1</i>	−0.10999201454742399	3785
<i>RP1 – 241P17.1</i>	−0.10999201454742399	173
<i>TTLL10</i>	0.109893101979087	842

<i>HLA – DPB1</i>	−0.109868885925718	2359
<i>C1QL3</i>	−0.10977522125084301	1005
<i>C3orf55</i>	−0.10963935043367801	70
<i>XXbac – BPG181M17.5</i>	−0.10942442445153699	62
<i>HCK</i>	−0.109007215333146	129
<i>PAH</i>	−0.108909828259471	3769
<i>DBX1</i>	−0.10859791759907	433
<i>FGF5</i>	−0.108480853512708	2173
<i>FOXE1</i>	−0.108460640192918	656
<i>GABRA3</i>	−0.108322288361311	1144
<i>THBS1</i>	−0.108012224702111	1356
<i>ANO7P1</i>	−0.10800653773188899	972
<i>DVL1</i>	0.107997667928319	2186
<i>AXIN2</i>	−0.10793463833573499	2520
<i>LRG_296</i>	−0.10793463833573499	2495
<i>RN7SL762P</i>	−0.107759788699141	2658
<i>RNU6 – 885P</i>	−0.107599168427578	4444
<i>RP1 – 170O19.14</i>	−0.10749113418684	654
<i>PDZRN3</i>	−0.10747876722784699	2713

<i>PDZRN3 – AS1</i>	−0.10747876722784699	512
<i>DLX5</i>	0.107441203156391	1927
<i>PDGFRA</i>	0.107429344813642	54
<i>LHFPL3 – AS1</i>	0.107334363324657	132
<i>RP5 – 1121A15.3</i>	0.10730918740216799	1255
<i>MIR3189</i>	0.107180741689341	228
<i>DLK1</i>	−0.10705750690709399	995
<i>RPP25</i>	0.10701489299339401	1718
<i>SCAMP5</i>	0.10701489299339401	1473
<i>RP11 – 85B7.2</i>	0.106864994393845	1445
<i>RP3 – 395M20.2</i>	0.106787124473451	1190
<i>RP3 – 395M20.3</i>	0.106787124473451	29
<i>KIF21B</i>	0.10670070046810499	371
<i>RP11 – 168O16.1</i>	0.10670070046810499	122
<i>MIR124 – 2</i>	−0.106642175623474	192
<i>OSR1</i>	0.106639719481052	3067
<i>MEOX2</i>	−0.106407197379582	41
<i>AC005550.4</i>	−0.106407197379582	1523

<i>MIR21</i>	−0.106337745838559	126
<i>PDPN</i>	−0.106242503306801	739
<i>CCDC64B</i>	0.106020719334687	1552
<i>CHAD</i>	−0.105780580885399	133
<i>MIR4297</i>	0.105659833057068	57
<i>PCDHGA11</i>	−0.105229756006415	166
<i>PCDHGB8P</i>	−0.105229756006415	4923
<i>KLHL14</i>	0.105227187967704	49
<i>SLC5A8</i>	−0.105161186803192	731
<i>ITGBL1</i>	0.105060953675732	473
<i>GHSR</i>	−0.104886809586504	3
<i>HOXA5</i>	0.104859995048034	742
<i>RP11 – 624C23.1</i>	−0.104854886848308	2550
<i>GS1 – 72M22.1</i>	−0.104854886848308	92
<i>NEFM</i>	−0.104854886848308	1611
<i>GPT2</i>	0.10480590801688799	821
<i>SFRP2</i>	−0.104518164900059	250

<i>RNU2 – 60P</i>	–0.10451297615451501	3723
<i>RGS7</i>	–0.104279302631394	877
<i>H2BFWT</i>	–0.104247652076186	49
<i>H2BFM</i>	–0.104247652076186	39
<i>ARMC4</i>	–0.104246364439679	214
<i>RP11 – 218D6.4</i>	–0.104246364439679	524
<i>CTSC</i>	0.104212710647739	1785
<i>LRG_50</i>	0.104212710647739	1771
<i>TSPYL5</i>	–0.10418878749230701	52
<i>WDR46</i>	0.10403223430929	1762
<i>PFDN6</i>	0.10403223430929	1537
<i>AC005559.3</i>	0.10400729134155599	4101
<i>GPRC5A</i>	0.103865906867918	11
<i>UBALD2</i>	–0.103836859804503	590
<i>RP11 – 285E9.6</i>	0.10382569025496	1550
<i>40057</i>	0.10382569025496	222
<i>TMEM261P1</i>	–0.10370450501645	106
<i>SNORD115 – 1</i>	–0.10370450501645	1153
<i>SNORD115 – 2</i>	–0.10370450501645	3065
<i>GHSR</i>	–0.10361076137165701	270
<i>GREB1</i>	–0.10301523610233899	1480
<i>RP1 – 40E16.2</i>	–0.102745208692728	2038
<i>RP11 – 513O17.2</i>	–0.102622883478428	864

<i>SLC7A2</i>	0.10257945756098601	616
<i>RP11 – 284F21.9</i>	0.102330099226989	1639
<i>BCAN</i>	0.102330099226989	1253
<i>RP11 – 284F21.10</i>	0.102330099226989	2242
<i>NR2F2 – AS1</i>	0.10229919224851899	1618
<i>NR2F2</i>	0.10229919224851899	195
<i>CTD – 2666L21.1</i>	−0.102230147125593	367
<i>CERKL</i>	−0.102125446159189	2158
<i>NEUROD1</i>	−0.102125446159189	2369
<i>AC013733.3</i>	−0.102125446159189	4606
<i>RP13 – 131K19.2</i>	−0.102058470880869	264
<i>RP13 – 131K19.7</i>	−0.102058470880869	3759
<i>P4HTM</i>	−0.102058470880869	162
<i>CNTN1</i>	0.101988875236978	1275
<i>RP11 – 32K4.1</i>	−0.10197902083079299	281
<i>LINC00966</i>	−0.10197902083079299	4487
<i>KCNV1</i>	−0.10189553138667699	4
<i>CRYBA2</i>	0.101866327505896	911
<i>UPP1</i>	−0.10179139338468	1678
<i>SRGN</i>	−0.101470034442445	511
<i>FOXE3</i>	−0.101353260878196	577
<i>AC099754.1</i>	−0.101207719948323	587
<i>LRRC3B</i>	−0.101207719948323	703

<i>IKBKG</i>	−0.101179944369073	1568
<i>FAM197Y1</i>	−0.101155812375696	892
<i>TSPY15P</i>	−0.101155812375696	130
<i>RGMA</i>	0.10108409025040201	974
<i>KCNH7</i>	−0.100983554882114	535
<i>OTX2</i>	−0.100920098537103	1229
<i>OTX2 – AS1</i>	−0.100920098537103	3933
<i>TMEM155</i>	−0.100674866753929	543
<i>AC079341.1</i>	−0.100674866753929	297
<i>RP11 – 609N14.1</i>	−0.100622143068377	2913
<i>SDR42E1</i>	−0.10056489087438	135
<i>OR2D2</i>	−0.100467407127736	185
<i>ZNF204P</i>	−0.100280738631745	3327
<i>ZNF391</i>	−0.100280738631745	237
<i>KRT79</i>	−0.10022075519222599	262
<i>RP11 – 153F5.3</i>	−0.10022075519222599	3370
<i>WT1</i>	−0.100119136435286	1537
<i>WT1 – AS</i>	−0.100119136435286	1649
<i>GRB7</i>	0.10004974772408801	216
<i>MSX1</i>	0.10003454190038	3886

Appendix D

Table of DMR's Functional Annotation (GO)

Supplementary Data File

Description: The accompanying table shows the functional enrichment of Gene Ontology biological processes of DMR's founded in melanoma biopsies pre and post MAPKi treatment.

<i>ID</i>	<i>Description</i>	<i>GeneRatio</i>
GO : 0007389	<i>patterns specification process</i>	18/116
GO : 0003002	<i>regionalization</i>	16/116
GO : 0009952	<i>anterior/posterior patterns specification</i>	13/116
GO : 0048705	<i>skeletal system morphogenesis</i>	13/116
GO : 0048736	<i>appendage development</i>	12/116
GO : 0060173	<i>limb development</i>	12/116
GO : 0030326	<i>embryonic limb morphogenesis</i>	10/116
GO : 0035113	<i>embryonic appendage morphogenesis</i>	10/116
GO : 0001501	<i>skeletal system development</i>	17/116
GO : 0035107	<i>appendage morphogenesis</i>	10/116
GO : 0035108	<i>limb morphogenesis</i>	10/116

GO : 0048562	<i>embryonic organ morphogenesis</i>	13/116
GO : 0048568	<i>embryonic organ development</i>	15/116
GO : 0023019	<i>signal transduction involved in regulation of gene expression</i>	5/116
GO : 0061448	<i>connective tissue development</i>	11/116
GO : 0060485	<i>mesenchyme development</i>	11/116
GO : 0035270	<i>endocrine system development</i>	8/116
GO : 0060021	<i>palate development</i>	7/116
GO : 0001837	<i>epithelial to mesenchymal transition</i>	8/116
GO : 0051216	<i>cartilage development</i>	9/116
GO : 0048762	<i>mesenchymal cell differentiation</i>	9/116
GO : 0090596	<i>sensory organ morphogenesis</i>	10/116
GO : 0042471	<i>ear morphogenesis</i>	7/116
GO : 0007548	<i>sex differentiation</i>	10/116
GO : 0045137	<i>development of primary sexual characteristics</i>	9/116
GO : 0035239	<i>tube morphogenesis</i>	11/116
GO : 0010717	<i>regulation of epithelial to mesenchymal transition</i>	6/116
GO : 0060349	<i>bone morphogenesis</i>	6/116
GO : 0060348	<i>bone development</i>	8/116
GO : 0007369	<i>gastrulation</i>	8/116
GO : 0048704	<i>embryonic skeletal system morphogenesis</i>	6/116

GO : 0060562	<i>epithelialtubemorphogenesis</i>	10/116
GO : 2000027	<i>regulationoforganmorphogenesis</i>	9/116
GO : 0045165	<i>cellfatecommitment</i>	9/116
GO : 0009954	<i>proximal/distalpatternformation</i>	4/116
GO : 0035051	<i>cardiocytedifferentiation</i>	7/116
GO : 0060537	<i>muscleissuedevelopment</i>	11/116
GO : 0035115	<i>embryonicforelimbmorphogenesis</i>	4/116
GO : 0042733	<i>embryonicdigitmorphogenesis</i>	5/116
GO : 0043583	<i>eardevelopment</i>	8/116
GO : 0008406	<i>gonaddevelopment</i>	8/116
GO : 0048608	<i>reproductivestructuredevelopment</i>	11/116
GO : 0035136	<i>forelimbmorphogenesis</i>	4/116
GO : 0061458	<i>reproductivesystemdevelopment</i>	11/116
GO : 0072148	<i>epithelialcellfatecommitment</i>	3/116
GO : 0048706	<i>embryonicskeletalsystemdevelopment</i>	6/116
GO : 0030900	<i>forebraindevelopment</i>	10/116
GO : 0072210	<i>metanephricnephrondevelopment</i>	4/116
GO : 0072224	<i>metanephricglomerulusdevelopment</i>	3/116
GO : 0040013	<i>negativeregulationoflocomotion</i>	9/116

ID	Official Gene Symbol
GO:0007389	LRP5/SFRP2/DBX1/AXIN2/ARMC4/NR2F2/WT1/MSX1
GO:0003002	OTX2/LRP5/SFRP2/MEOX2/GREM1/DBX1/AXIN2/NR2F2/NEUROD1/WT1/MSX1
GO:0009952	OTX2/SFRP2/MEOX2/AXIN2/OSR1/HOXA5/NR2F2/NEUROD1/WT1/MSX1
GO:0048705	GSC/SFRP2/IRX5/AXIN2/DLX5/PDGFR/CHAD/HOXA5/MSX1
GO:0048736	PITX1/HOXA10/SFRP2/MEOX2/GREM1/DLX5/OSR1/NR2F2/MSX1
GO:0060173	PITX1/LRP5/HOXA11/SFRP2/MEOX2/GREM1/DLX5/MSX1
GO:0030326	PITX1/LRP5/HOXA10/SFRP2/GREM1/DLX5/OSR1/MSX1
GO:0035113	PITX1/LRP5/HOXA11/SFRP2/GREM1/DLX5/OSR1/MSX1
GO:0035107	TBX5/PITX1/LRP5/HOXA10/SFRP2/GREM1/DLX5/OSR1/MSX1
GO:0035108	TBX5/PITX1/LRP5/HOXA11/SFRP2/GREM1/DLX5/OSR1/MSX1
GO:0048562	GSC/POU3F4/IRX5/FOXE1/DVL1/DLX5/PDGFR/OSR1/NEUROD1/MSX1
GO:0048568	GSC/ASCL2/POU3F4/IRX5/FOXE1/DVL1/DLX5/PDGFR/EUROD1/MSX1
GO:0023019	GSC/FGF5/PDGFR/NEUROD1/MSX1
GO:0061448	PITX1/LRP5/SFRP2/GREM1/AXIN2/OSR1/MIR21/WT1/MSX1
GO:0060485	SFRP2/PDPN/GREM1/AXIN2/OSR1/MIR21/HOXA5/WT1/MSX1
GO:0035270	PITX1/NR0B1/FOXE1/PDGFR/HOXA5/NEUROD1/WT1/MSX1
GO:0060021	PRDM16/MEOX2/FOXE1/DLX5/PDGFR/OSR1/MSX1
GO:0001837	GSC/TBX5/SFRP2/PDPN/GREM1/AXIN2/MIR21/MSX1
GO:0051216	PITX1/HOXA11/SFRP2/GREM1/AXIN2/OSR1/MIR21/MSX1
GO:0048762	GSC/TBX5/SFRP2/PDPN/GREM1/AXIN2/OSR1/MIR21/MSX1
GO:0090596	GSC/POU3F4/LRP5/NTN1/IRX5/DVL1/DLX5/OSR1/FOXE3/MSX1
GO:0042471	GSC/POU3F4/NTN1/DVL1/DLX5/OSR1/MSX1

GO:0007548	CDKL2/SFRP2/IRX5/NR0B1/PDGFRA/OSR1/WT1
GO:0045137	CBX2/SFRP2/IRX5/PDGFRA/OSR1/WT1
GO:0035239	LRP5/HOXA11/SFRP2/GREM1/DVL1/PDGFRA/MIR21/HOXA5/WT1
GO:0010717	TBX5/SFRP2/PDPN/GREM1/AXIN2/MIR21
GO:0060349	LRP5/HOXA11/SFRP2/AXIN2/DLX5/MSX1
GO:0060348	LRP5/HOXA11/SFRP2/GREM1/AXIN2/DLX5/CHAD/MSX1
GO:0007369	GSC/OTX2/LRP5/SFRP2/MIR145/DVL1/OSR1
GO:0048704	GSC/HOXA11/IRX5/PDGFRA/OSR1/HOXA5
GO:0060562	LRP5/HOXA11/SFRP2/GREM1/DVL1/OSR1/MIR21/WT1
GO:2000027	TBX5/GPC4/SFRP2/GREM1/DVL1/CHAD/WT1/MSX1
GO:0045165	GSC/PITX1/HOXA11/SFRP2/PDPN/DBX1/NR2F2/NEUROD1/WT1
GO:0009954	HOXA10/HOXA11/GREM1/OSR1
GO:0035051	TBX5/SIK1/MIR145/GREM1/PDGFRA/MIR21/WT1
GO:0060537	VGLL2/TBX5/PITX1/MIR145/GREM1/PDGFRA/OSR1/NR2F2/WT1
GO:0035115	TBX5/HOXA11/OSR1/MSX1
GO:0042733	LRP5/HOXA11/SFRP2/OSR1/MSX1
GO:0043583	GSC/POU3F4/NTN1/DVL1/DLX5/OSR1/NEUROD1/MSX1
GO:0008406	HOXA10/HOXA11/SFRP2/IRX5/NR0B1/PDGFRA/OSR1/WT1
GO:0048608	ASCL2/SFRP2/IRX5/NR0B1/PDGFRA/OSR1/GHSR/NR2F2/WT1
GO:0035136	TBX5/HOXA11/OSR1/MSX1
GO:0061458	ASCL2/SFRP2/IRX5/NR0B1/PDGFRA/OSR1/GHSR/NR2F2/WT1
GO:0072148	PDPN/NR2F2/NEUROD1
GO:0048706	GSC/HOXA11/IRX5/PDGFRA/OSR1/HOXA5
GO:0030900	POU3F4/OTX2/PITX1/DLX5/BCAN/NR2F2/NEUROD1/MSX1
GO:0072210	GREM1/PDGFRA/OSR1/WT1
GO:0072224	PDGFRA/OSR1/WT1
GO:0040013	TBX5/SFRP2/MEOX2/GREM1/THBS1/MIR124- 2/MIR21/GHSR/NR2F2

Bibliography

In:

Abdolmaleky, Hamid M, Sam Thiagalingam, and Marsha Wilcox (2005). “Genetics and epigenetics in major psychiatric disorders”. In: *American Journal of Pharmacogenomics* 5.3, pp. 149–160.

Afgan, Enis et al. (2018). “The Galaxy platform for accessible, reproducible and collaborative biomedical analyses 2018 update”. In: *Nucleic acids research* 46.W1, W537 W544 0305 1048.

Akalin A Kormaksson M, Li S et al (2012). “methylKit a comprehensive R package for the analysis of genome-wide DNA methylation profiles.” In: *Genome Biology* 13.

Anaconda, Inc. (2017). *Conda — Conda documentation*. <https://conda.io/docs/>. Accessed 8 Oct. 2018.

Antequera, Francisco, Donald Macleod, and Adrian P Bird (1989). “Specific protection of methylated CpGs in mammalian nuclei”. In: *Cell* 58.3, pp. 509–517.

Aryee, Martin J. et al. (2014). “Minfi a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays”. In: *Bioinformatics* 30.10, p. 1363 1369 1460 2059.

Ashburner, Michael et al. (2000). “Gene Ontology tool for the unification of biology”. In: *Nature genetics* 25.1, p. 25 1546 1718.

Assenov, Yassen et al. (2014). “Comprehensive analysis of DNA methylation data with RnBeads”. In: *Nature methods* 11.11, p. 1138 1548 7105.

Banks, Jo Ann and Nina Fedoroff (1989). “Patterns of developmental and heritable change in methylation of the Suppressor-mutator transposable element”. In: *Developmental genetics* 10.6, pp. 425–437.

- Biotechnology Information, U.S. National Library of Medicine National Center for (2018). *Published Tools*. <https://www.ncbi.nlm.nih.gov/pubmed/>.
- Bird, Adrian P. (1986). "CpGrich islands and the function of DNA methylation". In: *Nature* 321.6067, p. 209 213 0028 0836.
- Bird, Adrian P (1987). "CpG islands as gene markers in the vertebrate nucleus". In: *Trends in Genetics* 3, pp. 342–347.
- Bock, Christoph and Thomas Lengauer (2008). "Computational epigenetics". In: *Bioinformatics* 24.1, p. 1 10 1367 4803.
- Bolstad, Benjamin M et al. (2003). "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias". In: *Bioinformatics* 19.2, pp. 185–193.
- Boyle A. P., Guinney J. Crawford G. E. Furey T. S. (2008). "F-Seq a feature density estimator for high-throughput sequence tags." In: *Bioinformatics* 21.
- Brazma, Alvis et al. (2003). "ArrayExpress public repository for microarray gene expression data at the EBI". In: *Nucleic acids research* 31.1, p. 68 71 1362 4962.
- Breslow, Alexander (1970). "Thickness, cross sectional areas and depth of invasion in the prognosis of cutaneous melanoma". In: *Annals of surgery* 172.5, p. 902.
- Brown, Thomas C and Josef Jiricny (1988). "Different base/base mispairs are corrected with different efficiencies and specificities in monkey kidney cells". In: *Cell* 54.5, pp. 705–711.
- Busslinger, M, J Hurst, and RA Flavell (1983). "DNA methylation and the regulation of globin gene expression". In: *Cell* 34.1, pp. 197–206.
- Carbognin, Luisa et al. (2015). "Differential activity of nivolumab, pembrolizumab and MPDL3280A according to the tumor expression of programmed death ligand 1 (PD L1) sensitivity analysis of trials in melanoma, lung and genitourinary cancers". In: *PloS one* 10.6, e0130142 1932 6203.
- Carr, Ian M et al. (2007). "Sequence analysis and editing for bisulphite genomic sequencing projects". In: *Nucleic acids research* 35.10, e79.

- Cazaly, Emma et al. (2016). "Comparison of pre-processing methodologies for Illumina 450k methylation array data in familial analyses". In: *Clinical epigenetics* 8.1, p. 75.
- Clermont, P. L. et al. (2014). "Genotranscriptomic meta analysis of the Polycomb gene CBX2 in human cancers initial evidence of an oncogenic role". In: *British journal of cancer* 111.8, p. 1663 1532 1827.
- Consortium, ENCODE Project et al. (2004). "The ENCODE (ENCyclopedia of DNA elements) project". In: *Science* 306.5696, pp. 636–640.
- Consortium, Gene Ontology (2004). "The Gene Ontology (GO) database and informatics resource". In: *Nucleic acids research* 32.suppl_1, pp. D258–D261.
- Cooper, David and Hagop Youssoufian (1988). "The CpG dinucleotide and human genetic disease". In: *Human genetics* 78.2, p. 151 155.
- Daca-Roszak, Patrycja et al. (2015). "Impact of SNPs on methylation readouts by Illumina Infinium HumanMethylation450 BeadChip Array implications for comparative population studies". In: *BMC genomics* 16.1, p. 1003.
- Davis, Sean et al. (2012). "methyumi Handle Illumina methylation data". In: *R package version 2.0*.
- Dawson, Mark A and Tony Kouzarides (2012). "Cancer epigenetics from mechanism to therapy". In: *Cell* 150.1, pp. 12–27.
- Dedeurwaerder, Sarah et al. (2011). "Evaluation of the Infinium Methylation 450K technology". In: *Epigenomics* 3.6, p. 771 784 1750 1911.
- Dennis, Glynn et al. (2003). "DAVID database for annotation, visualization, and integrated discovery". In: *Genome biology* 4.9, R60 1474 760X.
- Developers, The Sean (2017). *Docker documentation*. <https://media.readthedocs.org/pdf/docker-sean/latest/docker-sean.pdf>. Accessed 8 Oct. 2018.
- Dhara, Sujoy K. and Steven L. Stice (2008). "Neural differentiation of human embryonic stem cells". In: *Journal of cellular biochemistry* 105.3, p. 633 640 0730 2312.

- Dorsky, Richard I., Randall T. Moon, and David W. Raible (1998). "Control of neural crest cell fate by the Wnt signalling pathway". In: *Nature* 396.6709, p. 370 1476 4687.
- Dummer, R. et al. (2012). "Cutaneous melanoma ESMO Clinical Practice Guidelines for diagnosis, treatment and follow up". In: *Annals of Oncology* 23, p. vii86 vii91.
- Dupin, Elisabeth and Nicole M Le Douarin (2003). "Development of melanocyte precursors from the vertebrate neural crest". In: *Oncogene* 22.20, p. 3016.
- Dweep, Harsh et al. (2011). "miRWalk database prediction of possible miRNA binding sites by "walking" the genes of three genomes". In: *Journal of biomedical informatics* 44.5, p. 839 847 1532 0464.
- Edgar, Ron, Michael Domrachev, and Alex E. Lash (2002). "Gene Expression Omnibus NCBI gene expression and hybridization array data repository". In: *Nucleic acids research* 30.1, p. 207 210 1362 4962.
- Egger, Gerda et al. (2004). "Epigenetics in human disease and prospects for epigenetic therapy". In: *Nature* 429.6990, p. 457 1476 4687.
- Erickson, Carol A. and Mark V. Reedy (1998). "5 Neural Crest Development The Interplay between Morphogenesis and Cell Differentiation". In: *Current topics in developmental biology*. Vol. 40. Elsevier, p. 177 209 0070 2153.
- Esteller, Manel and James G. Herman (2002). "Cancer as an epigenetic disease DNA methylation and chromatin alterations in human tumours". In: *The Journal of Pathology A Journal of the Pathological Society of Great Britain and Ireland* 196.1, p. 1 7 0022 3417.
- Feinberg, Andrew P. and Bert Vogelstein (1983). "A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity". In: *Analytical biochemistry* 132.1, p. 6 13 0003 2697.
- Fellows, Ian et al. (2012). "Deducer a data analysis GUI for R". In: *Journal of statistical Software* 49.8, pp. 1–15.

- Flaherty, Keith T (2012). “Dividing and conquering: controlling advanced melanoma by targeting oncogene-defined subsets”. In: *Clinical & experimental metastasis* 29.7, pp. 841–846.
- Fortin, Jean-Philippe et al. (2014). “Functional normalization of 450k methylation array data improves replication in large cancer studies”. In: *Genome biology* 15.11, p. 503.
- galaxyproject (2014). *Planemo documentation*. <https://planemo.readthedocs.io/en/latest/>. Accessed 8 Oct. 2018.
- Garraway, Levi A. et al. (2005). “Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma”. In: *Nature* 436.7047, p. 117 1476 4687.
- Gautam, Budhayash et al. (2018). “The Role of Bioinformatics in Epigenetics”. In: *Current trends in Bioinformatics: An Insight*. Springer, pp. 39–53.
- Gentleman, Robert C et al. (2004). “Bioconductor open software development for computational biology and bioinformatics”. In: *Genome biology* 5.10, R80.
- Ghiorzo, Paola et al. (1999). “Characterization of ligurian melanoma families and risk of occurrence of other neoplasia”. In: *International journal of cancer* 83.4, p. 441 448 0020 7136.
- Giebel, Lutz B. and Richard A. Spritz (1991). “Mutation of the KIT (mast/stem cell growth factor receptor) protooncogene in human piebaldism”. In: *Proceedings of the National Academy of Sciences* 88.19, p. 8696 8699 0027 8424.
- Gilbertson, Richard J. et al. (2006). “Mutational analysis of PDGFRRAS/MAPK pathway activation in childhood medulloblastoma”. In: *European journal of cancer* 42.5, p. 646 649 0959 8049.
- Glazar, Petar, Panagiotis Papavasileiou, and Nikolaus Rajewsky (2014). “circBase a database for circular RNAs”. In: *Rna* 20.11, p. 1666 1670 1355 8382.

- Goecks, Jeremy, Anton Nekrutenko, and James Taylor (2010). "Galaxy a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences". In: *Genome biology* 11.8, R86 1474 760X.
- Gordon, Randy (2013). "Skin cancer an overview of epidemiology and risk factors". In: vol. 29. Elsevier, p. 160 169 0749 2081.
- Gospodarowicz Mary, Brierley James and Christian Wittekind (2017). "TNM classification of malignant tumours". In: *John Wiley and Sons*.
- Griffiths Jones, Sam et al. (2006). "miRBase microRNA sequences, targets and gene nomenclature". In: *Nucleic acids research* 34.suppl 1, p. D140 D144 1362 4962.
- Hackenberg, Michael, Guillermo Barturen, and Jose L. Oliver (2010). "NGSmethDB a database for next generation sequencing single cytosine resolution DNA methylation data". In: *Nucleic acids research* 39.suppl 1, p. D75 D79 0305 1048.
- Hansen, Kasper Daniel and Martin Aryee (2012). "minfi Analyze Illumina's 450k methylation arrays". In: *R package version* 1.0.
- Harbour, William, Roberson Elisha Duan Shenghui Cao Li Worley Lori Onken Michael, and Matatall Katie Helms Cynthia Bowcock Anne Council Laurin (2010). "Frequent mutation of BAP1 in metastasizing uveal melanomas". In: *Science* 330.6009, pp. 1410–1413.
- Harris, R Alan et al. (2010). "Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications". In: *Nature biotechnology* 28.10, p. 1097.
- Hauschild, Axel et al. (2012). "Dabrafenib in BRAF-mutated metastatic melanoma: a multicentre, open-label, phase 3 randomised controlled trial". In: *The Lancet* 380.9839, pp. 358–365.
- He, Ximiao et al. (2007). "MethyCancer the database of human DNA methylation and cancer". In: *Nucleic acids research* 36.suppl 1, p. D836 D841 0305 1048.

- Holliday, Robin (2006). "Epigenetics a historical overview". In: *Epigenetics* 1.2, p. 76 80 1559 2294.
- Hugo, Willy et al. (2015). "Non genomic and immune evolution of melanoma acquiring MAPKi resistance". In: *Cell* 162.6, p. 1271 1285 0092 8674.
- Humburg, Peter et al. (2011). "ChIPseqR analysis of ChIP seq experiments". In: *BMC bioinformatics* 12.1, p. 39 1471 2105.
- Illumina, Inc. (2018). *Infinium Methylation Assay Overview*. <https://emea.illumina.com/science/technology/beadarray-technology/infinium-methylation-assay.html>. Accessed 8 Oct. 2018.
- Jaffe, Andrew E. et al. (2012). "Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies". In: *International journal of epidemiology* 41.1, p. 200 209 1464 3685.
- Jakovcevski, Mira and Schahram Akbarian (2012). "Epigenetic mechanisms in neurological disease". In: *Nature medicine* 18.8, p. 1194.
- Jin, Eun Jung et al. (2001). "Wnt and BMP signaling govern lineage segregation of melanocytes in the avian embryo". In: *Developmental biology* 233.1, p. 22 37 0012 1606.
- Kanehisa, Minoru and Susumu Goto (2000). "KEGG kyoto encyclopedia of genes and genomes". In: *Nucleic acids research* 28.1, p. 27 30 1362 4962.
- Kawakami, Akinori and David E. Fisher (2017). "The master role of microphthalmia associated transcription factor in melanocyte and melanoma biology". In: *Laboratory Investigation* 97.6, p. 649 1530 0307.
- Kent, W James et al. (2002). "The human genome browser at UCSC". In: *Genome research* 12.6, pp. 996–1006.
- Khare, Satyajeet P. et al. (2011). "Histone relational knowledgebase of human histone proteins and histone modifying enzymes". In: *Nucleic acids research* 40.D1, p. D337 D342 1362 4962.
- Kishore, Kamal et al. (2015). "methylPipe and compEpiTools a suite of R packages for the integrative analysis of epigenomics data". In: *BMC bioinformatics* 16.1, p. 313 1471 2105.

- Kléber, Maurice et al. (2005). "Neural crest stem cell maintenance by combinatorial Wnt and BMP signaling". In: *The Journal of cell biology* 169.2, pp. 309–320.
- Klose, Robert J. and Adrian P. Bird (2006). "Genomic DNA methylation the mark and its mediators". In: *Trends in biochemical sciences* 31.2, p. 89 97 0968 0004.
- Kristensen, Lasse Sommer and Lise Lotte Hansen (2009). "PCR based methods for detecting single locus DNA methylation biomarkers in cancer diagnostics, prognostics, and response to treatment". In: *Clinical chemistry* 55.8, p. 1471 1483 0009 9147.
- Lawrence, Michael, Robert Gentleman, and Vincent Carey (2009). "rtracklayer an R package for interfacing with genome browsers". In: *Bioinformatics* 25.14, pp. 1841–1842.
- Lewis, Joe and Adrian Bird (1991). "DNA methylation and chromatin structure". In: *Febs Letters* 285.2, pp. 155–159.
- Love, Michael I, Wolfgang Huber, and Simon Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". In: *Genome biology* 15.12, p. 550.
- Lv, Jie et al. (2011). "DiseaseMeth a human disease methylation database". In: *Nucleic acids research* 40.D1, p. D1030 D1035 1362 4962.
- Maat, Willem et al. (2007). "Epigenetic inactivation of RASSF1a in uveal melanoma". In: *Investigative ophthalmology & visual science* 48.2, pp. 486–490.
- Manchester, The University of and University of Southampton (2018). *MyExperiment*. <http://www.myexperiment.org/>. Accessed 8 Oct. 2018.
- Marabita, Francesco et al. (2013a). "An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 Bead-Chip platform". In: *Epigenetics* 8.3, p. 333 346 1559 2294.
- Marabita, Francesco et al. (2013b). "An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 Bead-Chip platform". In: *Epigenetics* 8.3, pp. 333–346.

- Matloff, Norman (2011). *The art of R programming A tour of statistical software design*. No Starch Press.
- Meda, Francesca et al. (2011). "The epigenetics of autoimmunity". In: *Cellular & molecular immunology* 8.3, p. 226.
- Mitra A, Song J (2012). "WaveSeq A Novel Data-Driven Method of Detecting Histone Modification Enrichments Using Wavelets." In: *Bioinformatics* 9.
- Morris, Tiffany J. et al. (2013). "ChAMP 450k chip analysis methylation pipeline". In: *Bioinformatics* 30.3, p. 428 430 1460 2059.
- Nakabayashi, Kazuhiko (2017). "Illumina HumanMethylation BeadChip for Genome-Wide DNA Methylation Profiling Advantages and Limitations". In: *Handbook of Nutrition, Diet, and Epigenetics*. Springer, pp. 1–15.
- Nanda, Jagpreet Singh, Rahul Kumar, and Gajendra P. S. Raghava (2016). "dbEM A database of epigenetic modifiers curated from cancerous and normal genomes". In: *Scientific reports* 6, p. 19340 2045 2322.
- Niller, Hans Helmut, Hans Wolf, and Janos Minarovits (2009). "Epigenetic dysregulation of the host cell genome in Epstein–Barr virus-associated neoplasia". In: *Seminars in cancer biology*. Vol. 19. 3. Elsevier, pp. 158–164.
- Osaki, Mitsuhiro et al. (2013). "Decreased PITX1 gene expression in human cutaneous malignant melanoma and its clinicopathological significance". In: *European Journal of Dermatology* 23.3, p. 344 349 1167 1122.
- Paradis, Emmanuel, Julien Claude, and Korbinian Strimmer (2004). "APE analyses of phylogenetics and evolution in R language". In: *Bioinformatics* 20.2, pp. 289–290.
- Pepke, Shirley, Barbara Wold, and Ali Mortazavi (2009). "Computation for ChIP-seq and RNA-seq studies". In: *Nature methods* 6.11s, S22.
- Peters, T. J. et al. (2014). "DMRcate Illumina 450 K methylation array apatial analysis methods". In: *R package version* 1.0.
- Pfeifer, GP et al. (1990). "Polymerase chain reaction-aided genomic sequencing of an X chromosome-linked CpG island methylation patterns suggest clonal inheritance, CpG site autonomy, and an explanation of activity state

- stability.” In: *Proceedings of the National Academy of Sciences* 87.21, pp. 8252–8256.
- Pidsley, Ruth et al. (2013). “A data driven approach to preprocessing Illumina 450K methylation array data”. In: *BMC genomics* 14.1, p. 293 1471 2164.
- Poterlowicz K., Murat K. (2018a). *Published Tools*. https://testtoolshed.g2.bx.psu.edu/repository/browse_repositories_in_categorysort=name&operation=repositories_by_user&id=0a77a6371a54a53. Accessed 8 Oct. 2018.
- (2018b). *Source Code of EWAS Tools*. <https://github.com/kpbioteam>. Accessed 8 Oct. 2018.
- Pramio, Dimitrius T et al. (2017). “Epigenetic signature of differentially methylated genes in cutaneous melanoma”. In: *Applied Cancer Research* 37.1, p. 34.
- Racine, Jeffrey S (2012). “RStudio A Platform-Independent IDE for R and Sweave”. In: *Journal of Applied Econometrics* 27.1, pp. 167–172.
- Rakyan, Vardhman K. et al. (2011). “Epigenome wide association studies for common human diseases”. In: *Nature Reviews Genetics* 12.8, p. 529 1471 0064.
- Reik, Wolf et al. (1987). “Genomic imprinting determines methylation of parental alleles in transgenic mice”. In: *Nature* 328.6127, p. 248.
- Ritchie ME Phipson B, Wu D et al. (2015). “limma powers differential expression analyses for RNA-sequencing and microarray studies.” In: *Nucleic Acids Research* 2.
- Robinson, Mark D., Davis J. McCarthy, and Gordon K. Smyth (2010). “edgeR a Bioconductor package for differential expression analysis of digital gene expression data”. In: *Bioinformatics* 26.1, p. 139 140 1367 4803.
- Rozowsky, Joel et al. (2009). “PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls”. In: *Nature biotechnology* 27.1, p. 66.
- Sager, Ruth (1989). “Tumor suppressor genes: the puzzle and the promise”. In: *Science* 246.4936, pp. 1406–1412.

- Sandoval, Juan et al. (2011). "Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome". In: *Epigenetics* 6.6, p. 692 702 1559 2294.
- Scharpf RB Parmigiani G, Pevsner J Ruczinski I. (2008). "Hidden Markov models for the assessment of chromosomal alterations using high-throughput SNP arrays." In: *The annals of applied statistics* 2.
- Smittenaar, CR et al. (2016). "Cancer incidence and mortality projections in the UK until 2035". In: *British journal of cancer* 115.9, p. 1147.
- Spyrou, Christiana et al. (2009). "BayesPeak Bayesian analysis of ChIP seq data". In: *BMC bioinformatics* 10.1, p. 299 1471 2105.
- Sun, Guannan et al. (2013). "Statistical analysis of ChIP seq data with MO-SAICS". In: *Deep Sequencing Data Analysis*. Springer, p. 193 212.
- Teng, Li et al. (2015). "4DGenome a comprehensive database of chromatin interactions". In: *Bioinformatics* 31.15, p. 2560 2564 1460 2059.
- Tobin, Desmond J. (2006). "Biochemistry of human skin our brain on the outside". In: *Chemical Society Reviews* 35.1, p. 52 67.
- Tomczak Czerwinska, Wiznerowicz (2015). "The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge". In: *Contemporary oncology* 19.1A, A68.
- Uddin, Monica et al. (2010). "Epigenetic and immune function profiles associated with posttraumatic stress disorder". In: *Proceedings of the National Academy of Sciences*, p. 200910794.
- Uong, Audrey and Leonard I. Zon (2010). "Melanocytes in development and cancer". In: *Journal of cellular physiology* 222.1, p. 38 41 0021 9541.
- Urich, Mark A et al. (2015). "MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing". In: *Nature protocols* 10.3, p. 475.
- Whelan Alison, Bartsch Detlef Goodfellow Paul (1995). "A familial syndrome of pancreatic cancer and melanoma with a mutation in the CDKN2 tumor-suppressor gene". In: *New England Journal of Medicine* 333.15, pp. 975–977.

- Williams, Graham J et al. (2009). "Rattle a data mining GUI for R". In: *The R Journal* 1.2, pp. 45–55.
- Xin, Yurong et al. (2011). "MethylomeDB a database of DNA methylation profiles of the brain". In: *Nucleic acids research* 40.D1, p. D1245 D1249 1362 4962.
- Zhang, Yong et al. (2008). "Model based analysis of ChIP Seq (MACS)". In: *Genome biology* 9.9, R137 1474 760X.